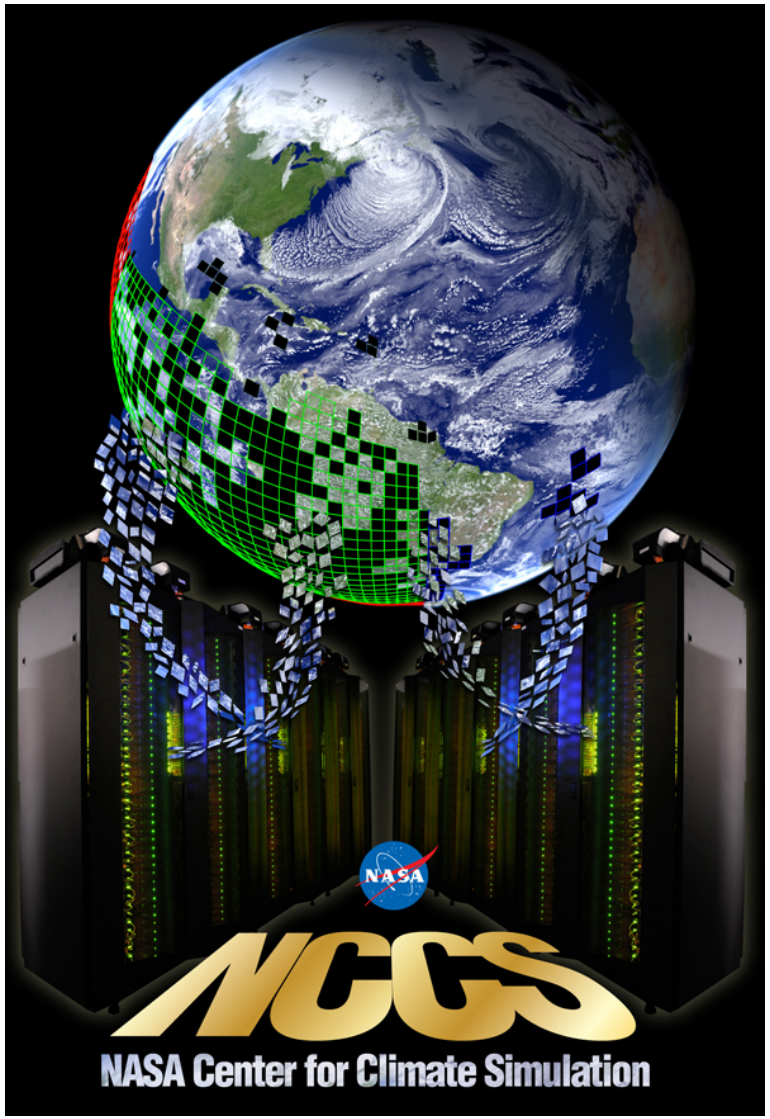
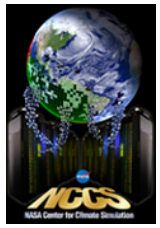


NCCS User Forum

December 7, 2010

Agenda – December 7, 2010



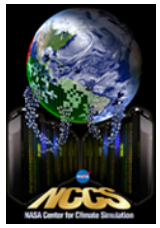
- **<u>Welcome & Introduction (Phil Webster, CISTO Chief)</u>**
- Current System Status (Fred Reitz, NCCS Operations Manager)
- SCU7 and Other NCCS Systems Updates (Dan Duffy, NCCS Lead Architect)
- Analysis Software Update (Tom Maxwell, NCCS Analysis Lead)
- User Services Update (Tyler Simon, NCCS User Services Group)
- Questions & Wrap-Up (Phil Webster)

Accomplishments

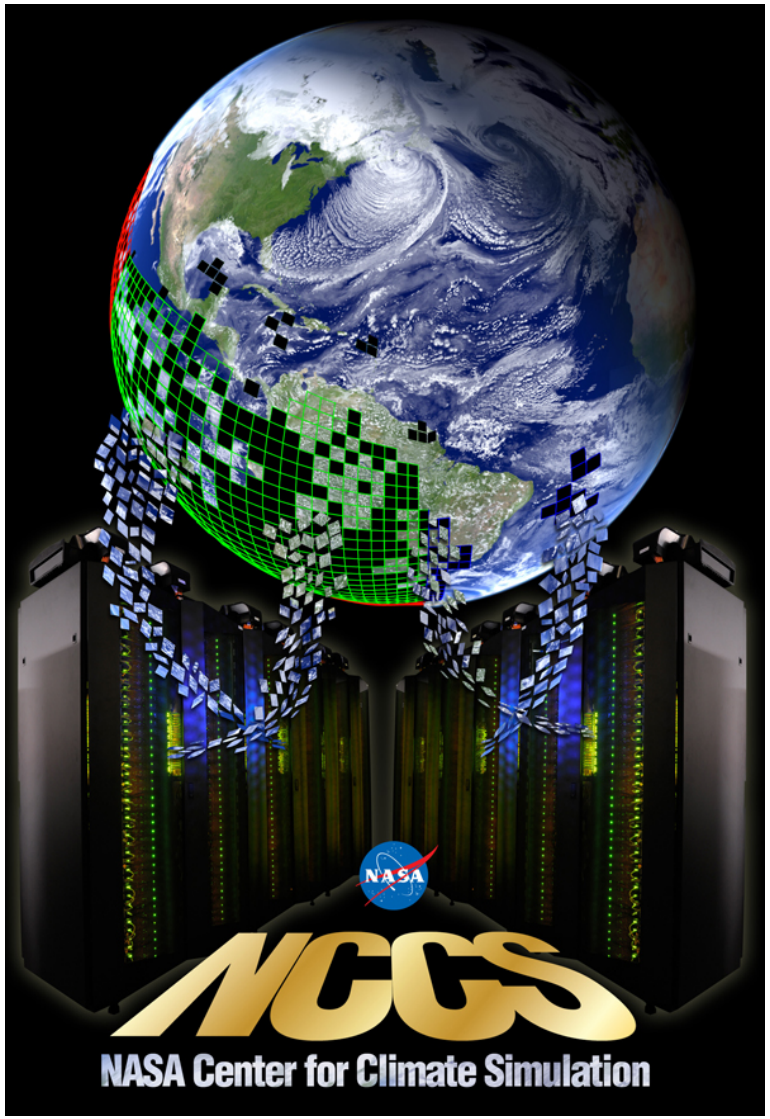
- Discover Linux cluster: SCU7 coming on line (159 TFLOPs peak)
  - Power and cooling issues being addressed by Dell at high levels

- Dirac Mass Storage archive (DMF)
  - Disk cache nearly quadrupled, to 480 TB
  - Server moved to Distributed DMF cluster
- Science support
  - GRIP field campaign
    - Genesis and Rapid Intensification Processes, completed September 2010
    - Provided monitoring, troubleshooting for timely execution of jobs
    - Supported forecast team via image and data download services on Data Portal
  - IPCC AR5 (ongoing)
    - Intergovernmental Panel on Climate Change – Fifth Reassessment
    - Climate jobs running on Discover
    - Data publication via Earth System Grid "Data Node" on DataPortal

Agenda – December 7, 2010



- Welcome & Introduction (Phil Webster, CISTO Chief)
- **Current System Status (Fred Reitz, NCCS Operations Manager)**
- SCU7 and Other NCCS Systems Updates (Dan Duffy, NCCS Lead Architect)
- Analysis Software Update (Tom Maxwell, NCCS Analysis Lead)
- User Services Update (Tyler Simon, NCCS User Services Group)
- Questions & Wrap-Up (Phil Webster)



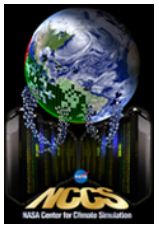
Operations and Maintenance

NASA Center for Climate Simulation (NCCS) Project

Fred Reitz

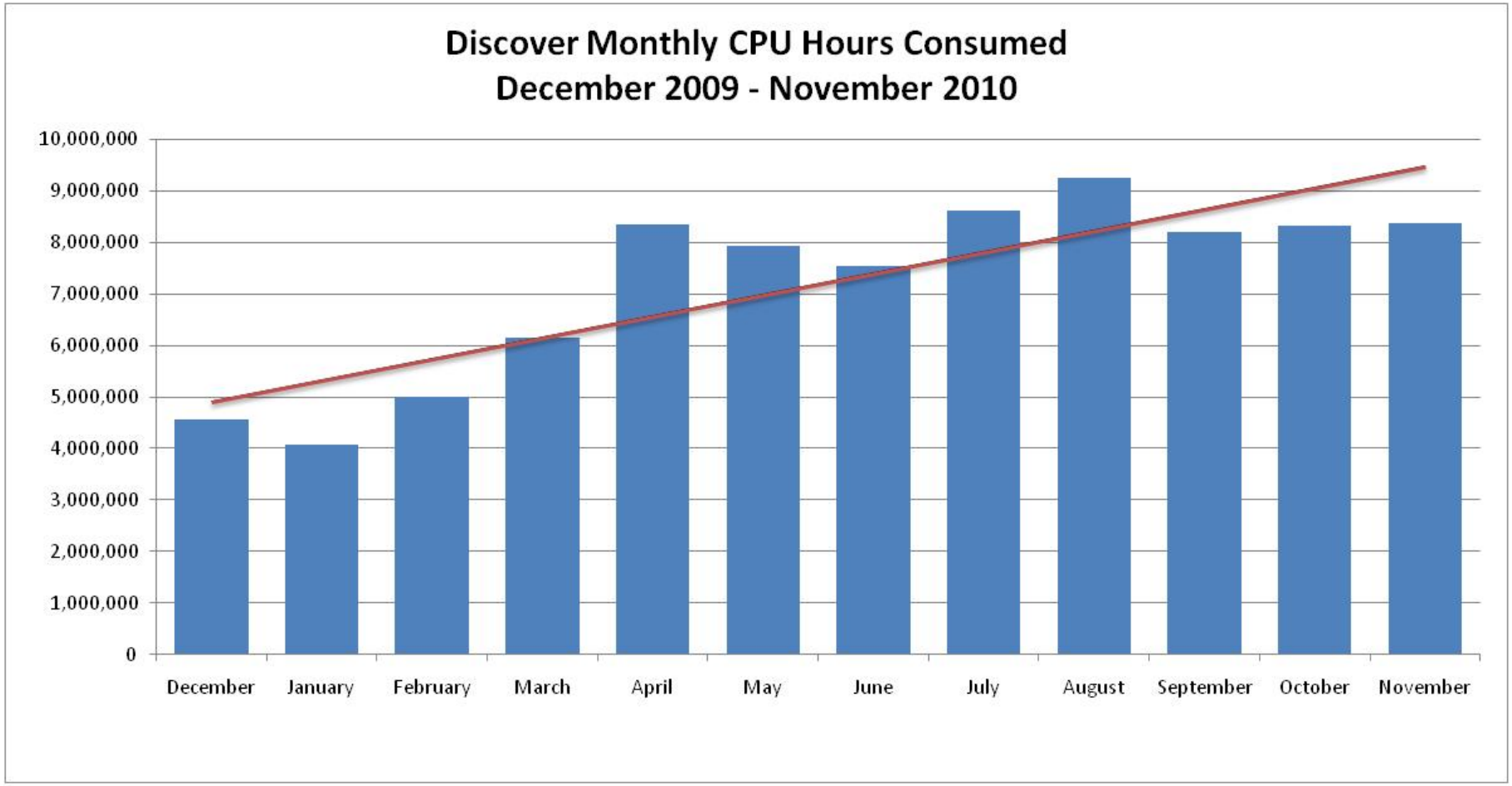
December 7, 2010

Accomplishments

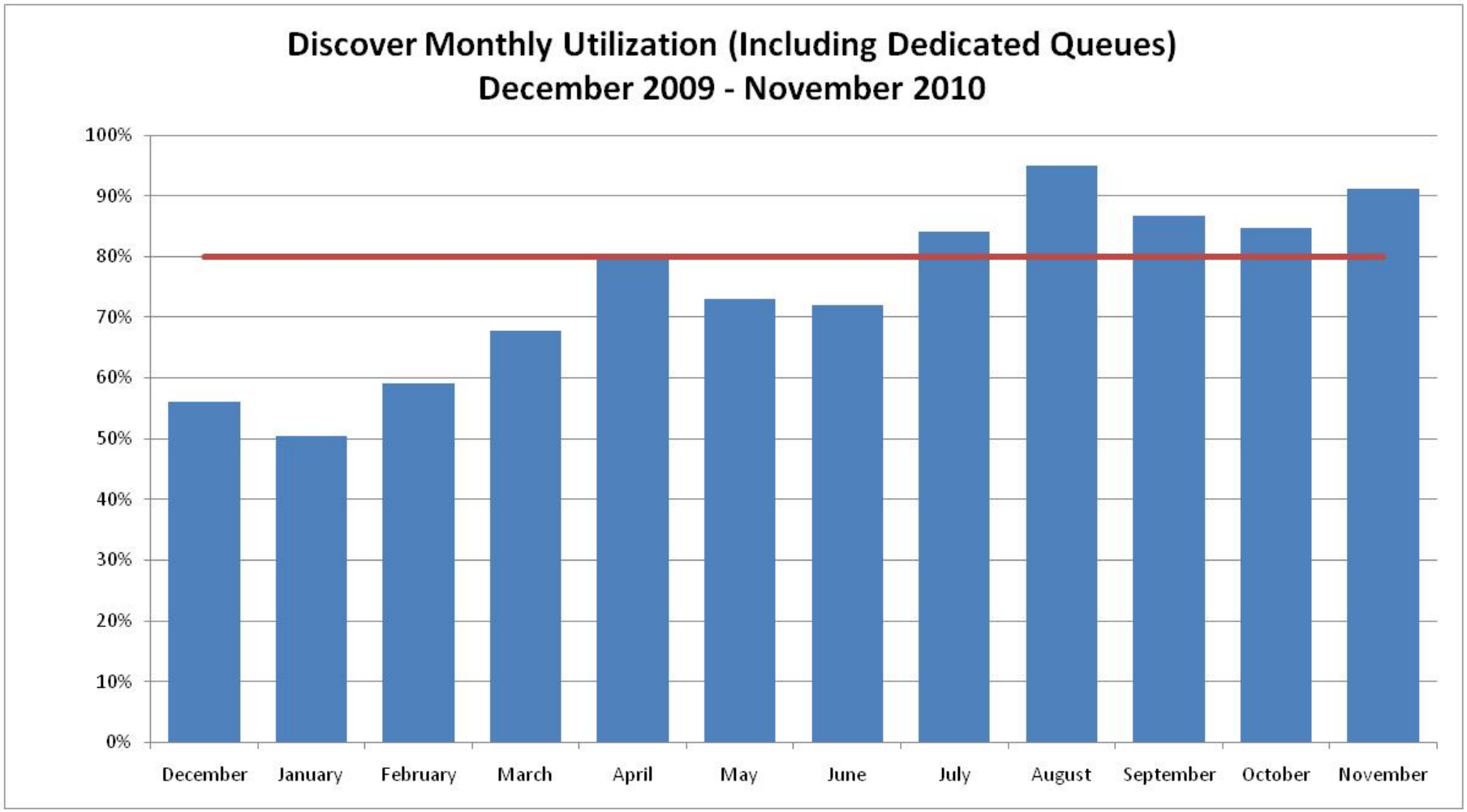


- *Discover*
  - Completed GRIP mission support
  - Continued SCU7 preparations
    - Installed SCU7 Intel Xeon "Westmere" hardware (159 TF peak)
    - Installed system image (SLES 11)
    - Integrated SCU7 into existing InfiniBand fabric with SCU5, SCU6
    - Troubleshooting power, cooling issues
- Mass Storage
  - Archive upgrade – placed new hardware into production
    - New machines, fast processors
    - Interactive logins: Two Nehalem servers
    - Parallel Data Movers serving NCCS'tape drives: Three Nehalem servers
    - DMF Daemon processes: Two Westmere servers in a high-availability configuration
    - NFS-serving to Discover: Two Nehalem servers
    - Scalable
    - Palm was too expensive to maintain
    - Increased disk storage
  - Troubleshooting issues
    - One of the first deployed in production, has the most demanding workload to date
- *DataPortal*
  - New database service (upgraded hardware)
  - Database service accessible via *Dali*, *Discover* login nodes
  - Additional storage (including *datastage*)

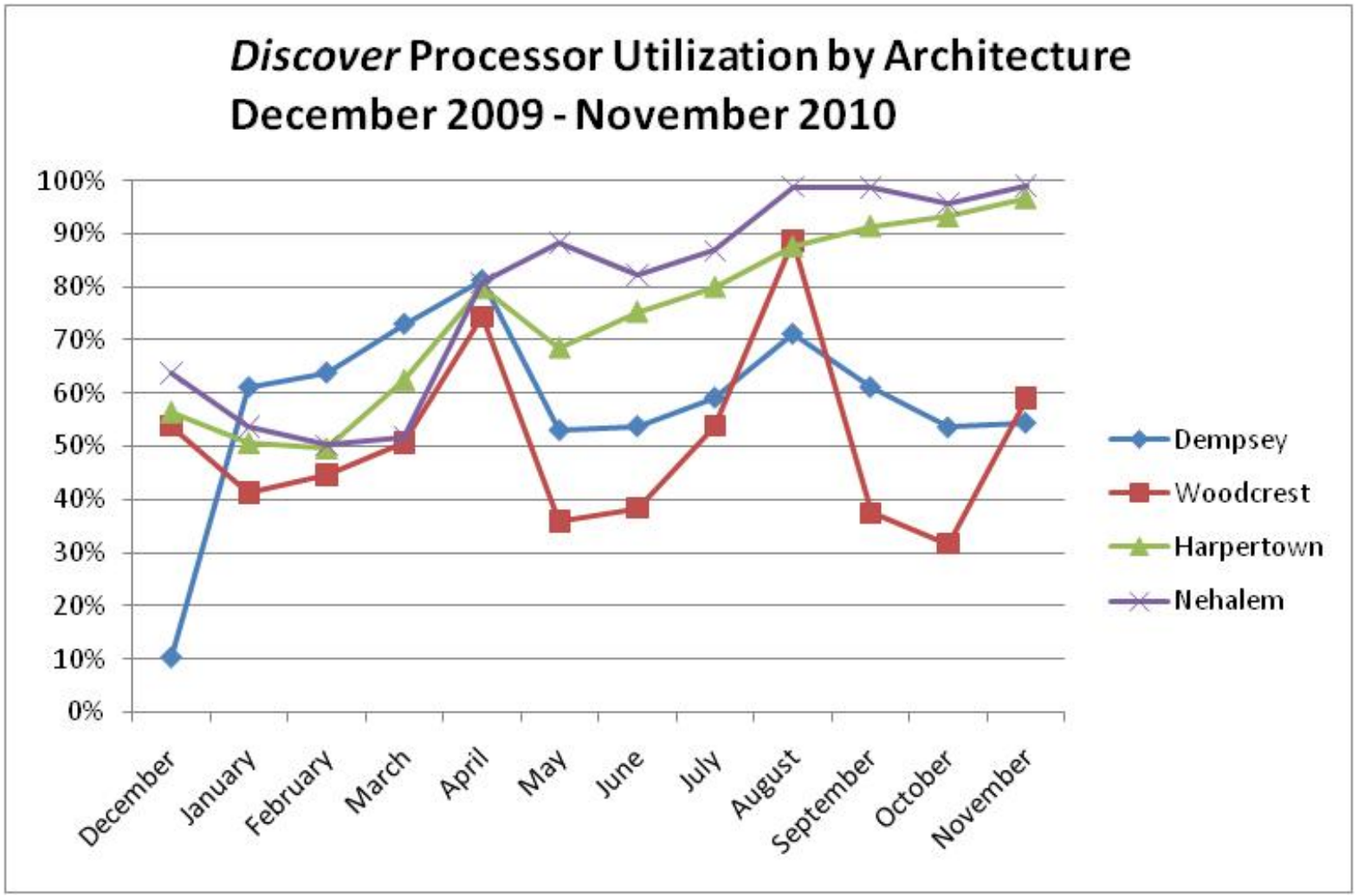
Discover Total CPU Consumption Past 12 Months (CPU Hours)



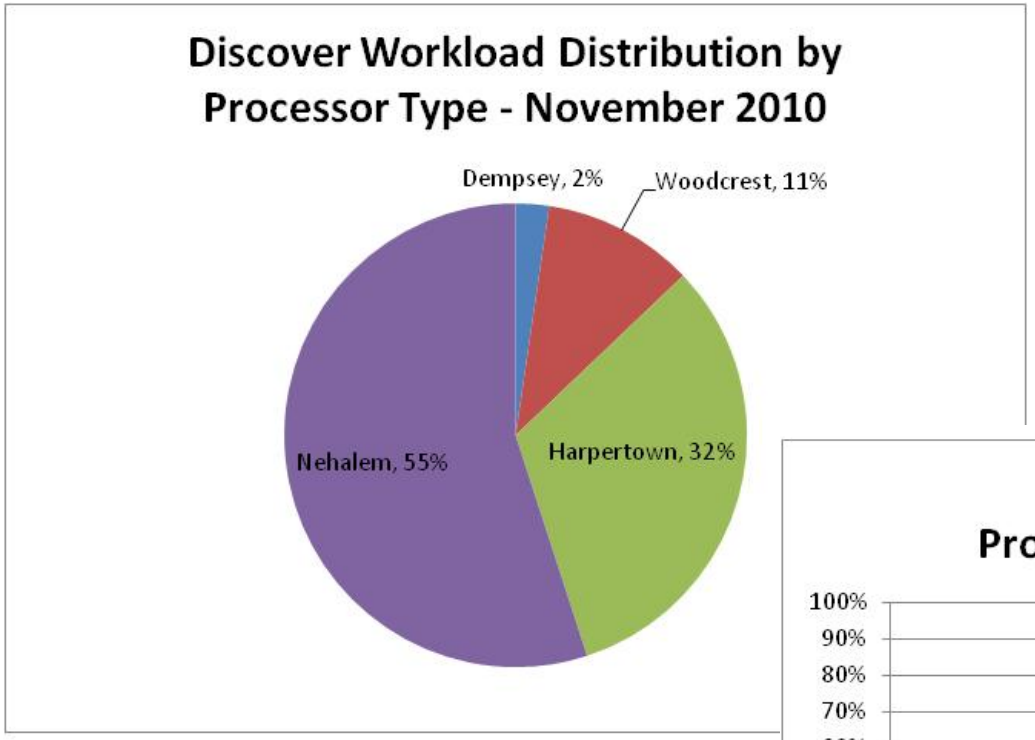
Discover Total Utilization
Past 12 Months (Percentage)

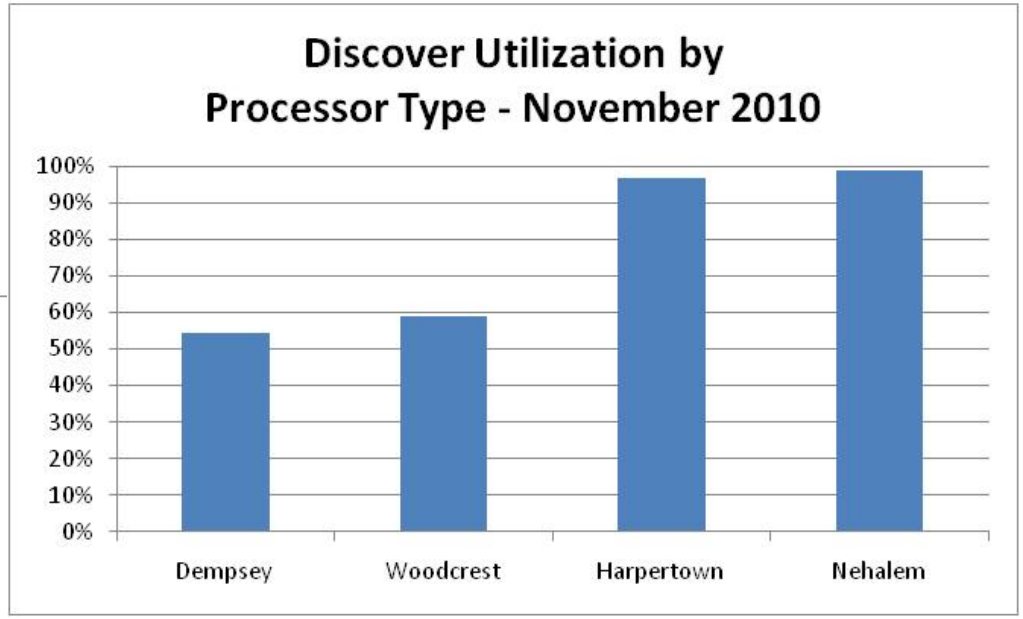


Discover Utilization by Architecture Past 12 Months

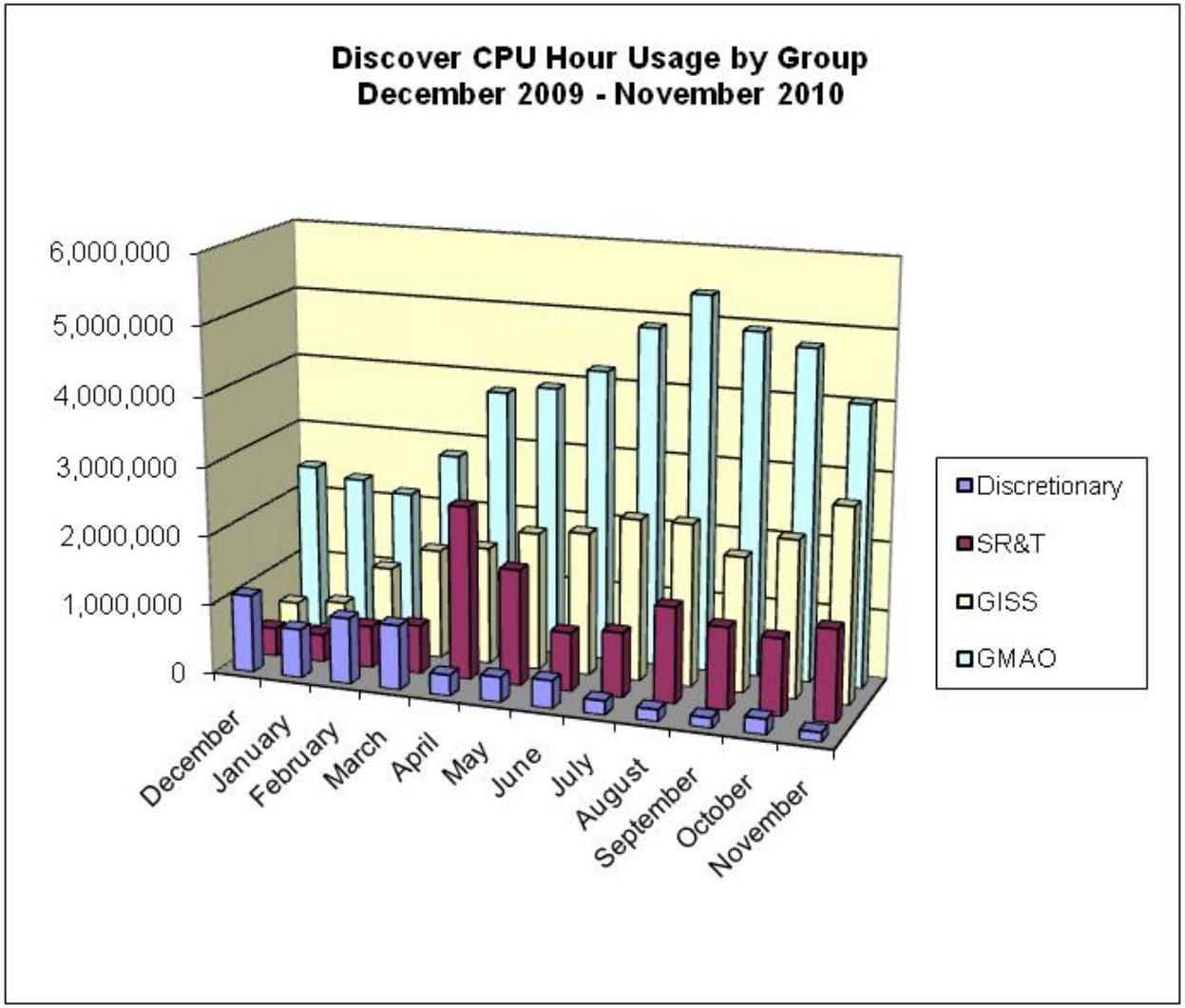


Discover Workload Distribution and Utilization by Architecture – October 2010

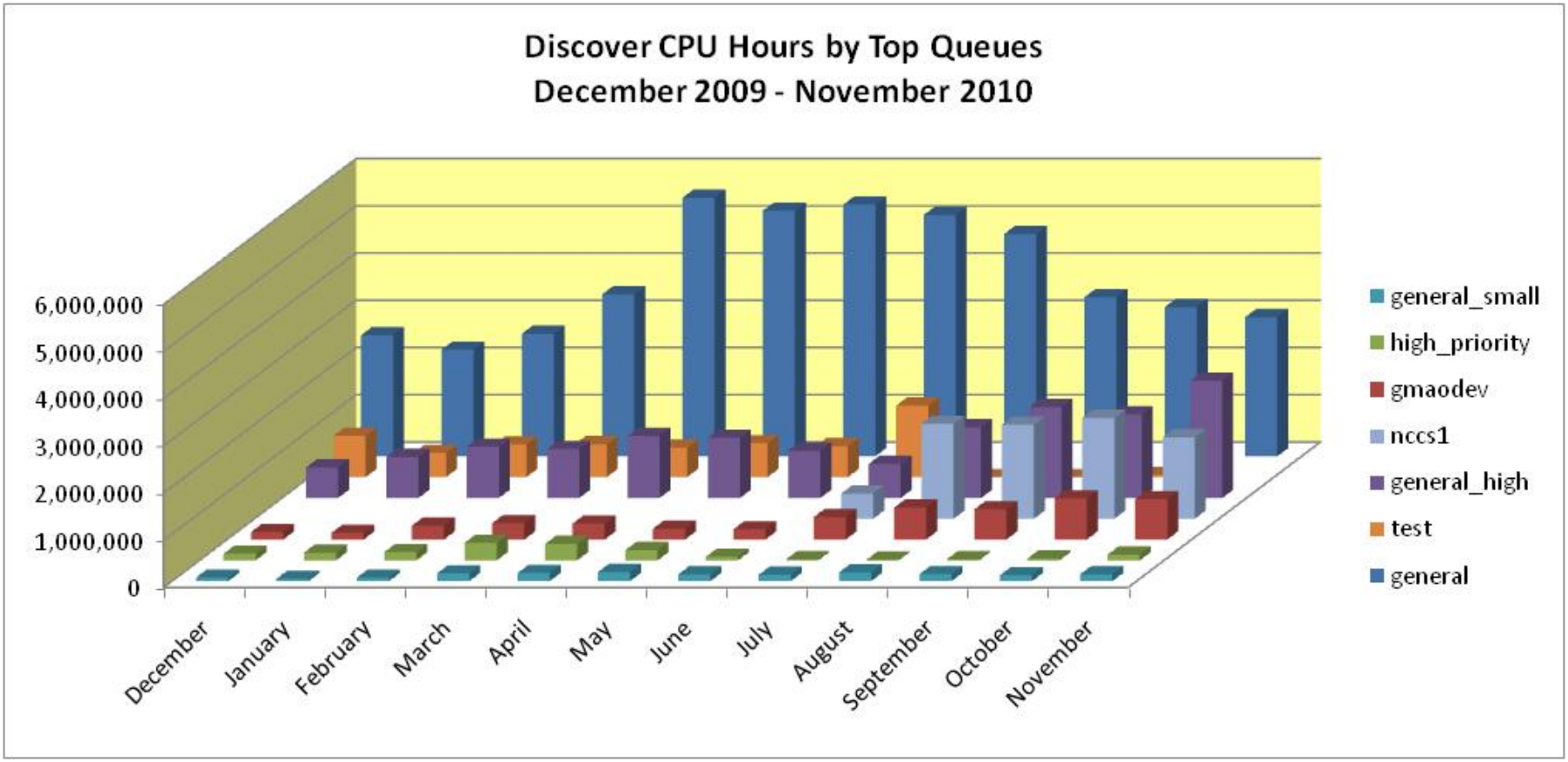




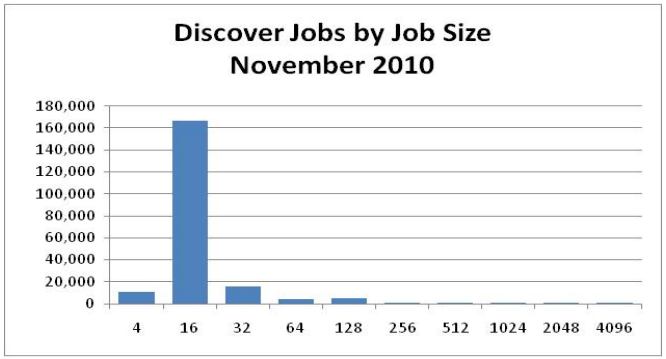
Discover CPU Consumption by Group Past 12 Months

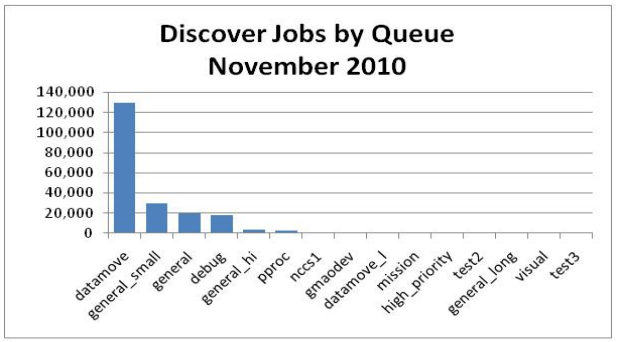


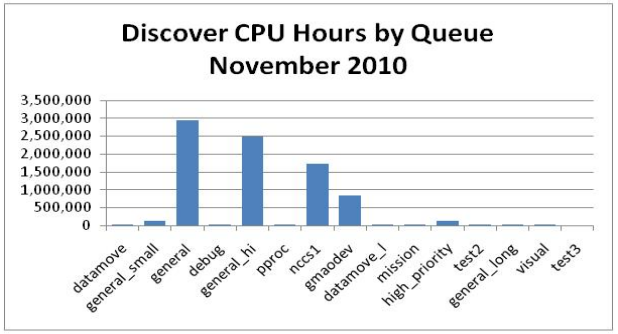
Discover CPU Consumption by Top Queues Past 12 Months

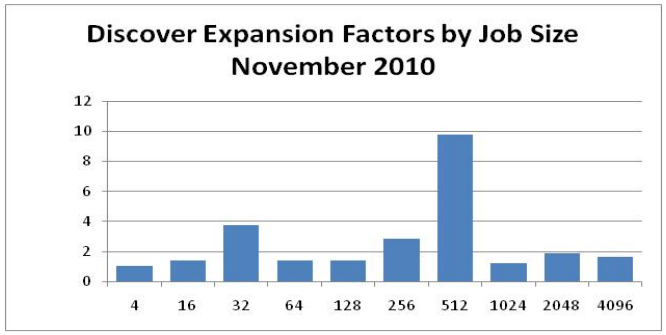


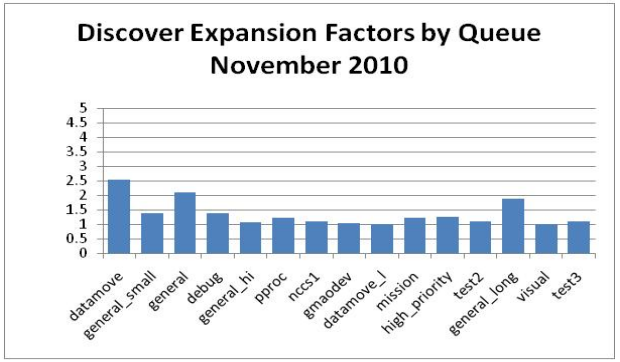
Discover Job Analysis by Job Size and Queue October 2010



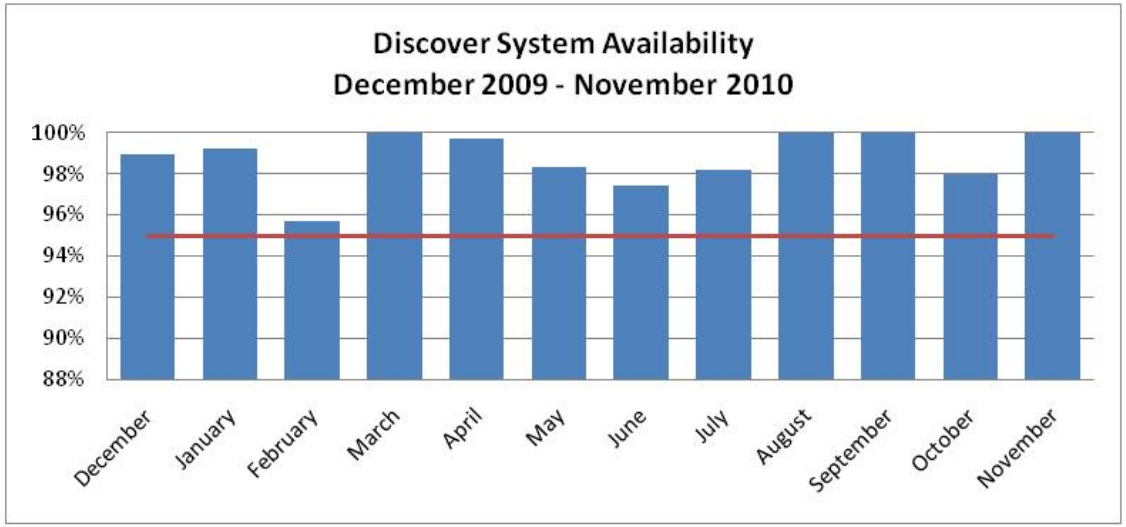


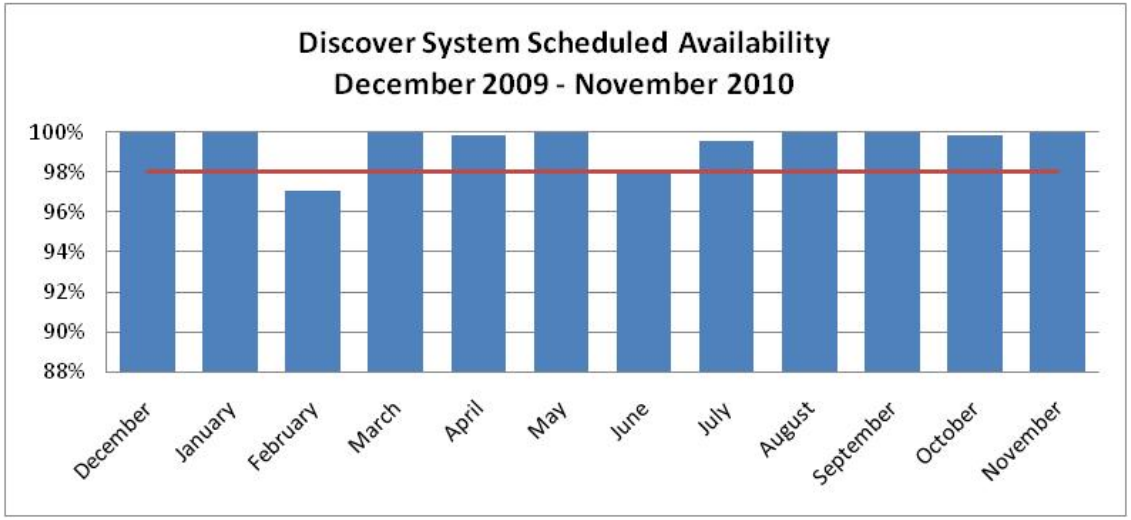




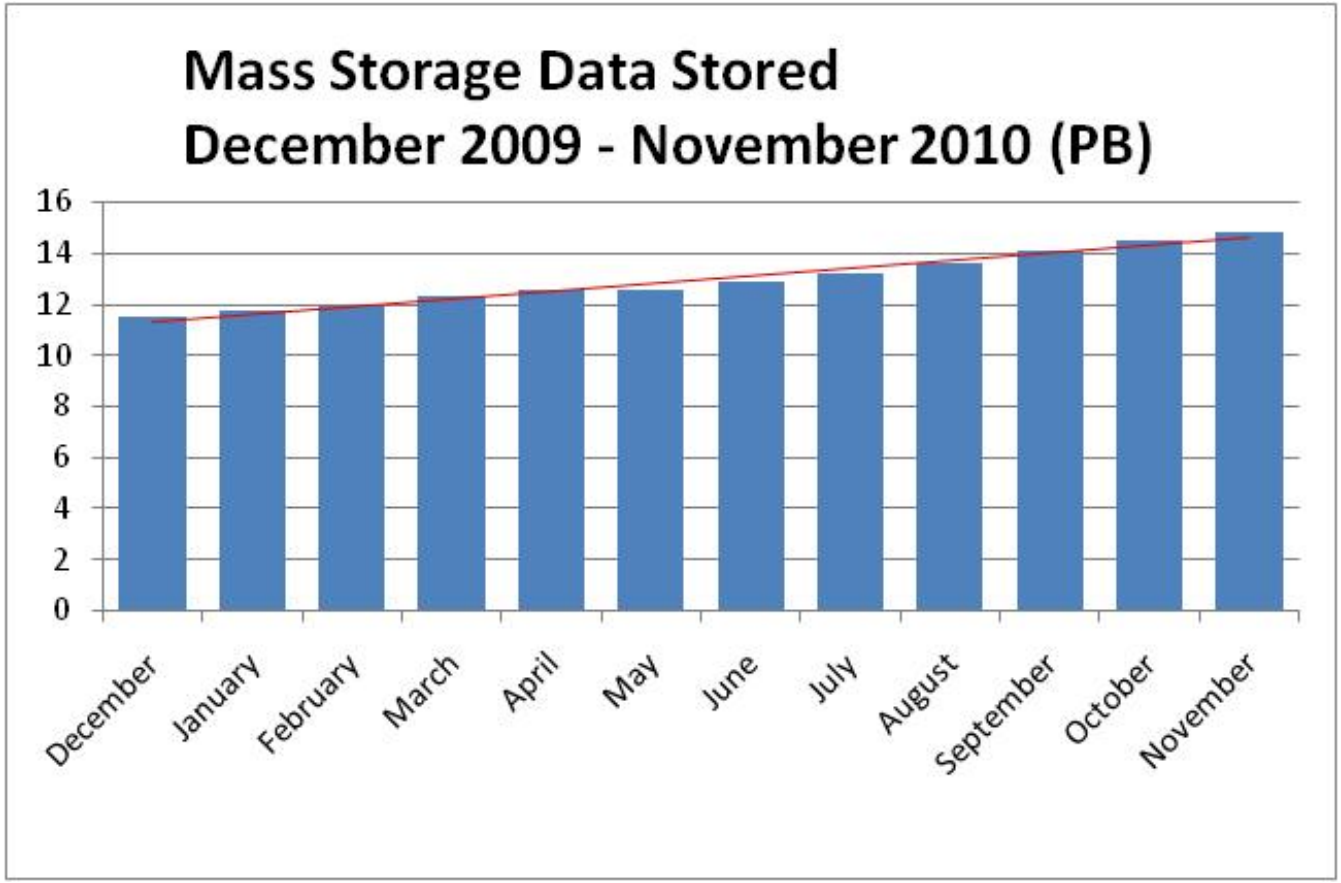


Discover Availability
Past 12 Months



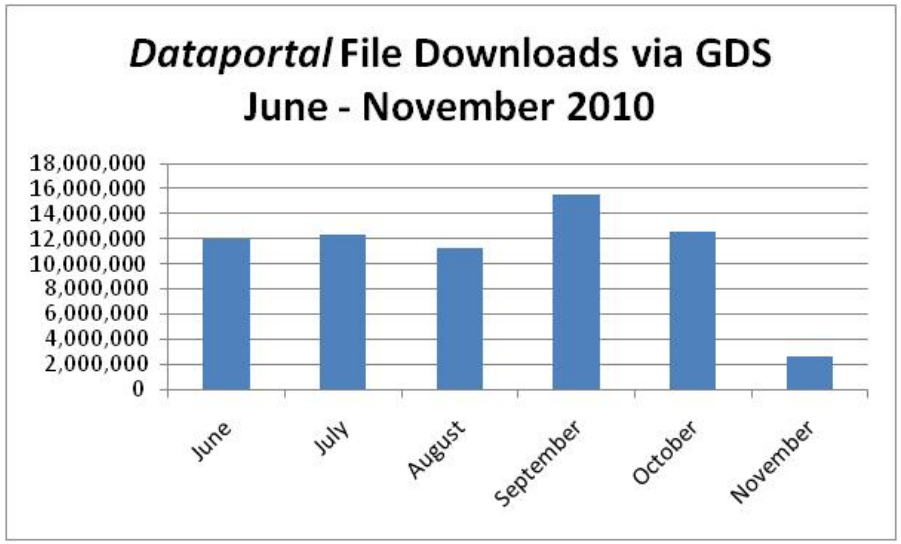


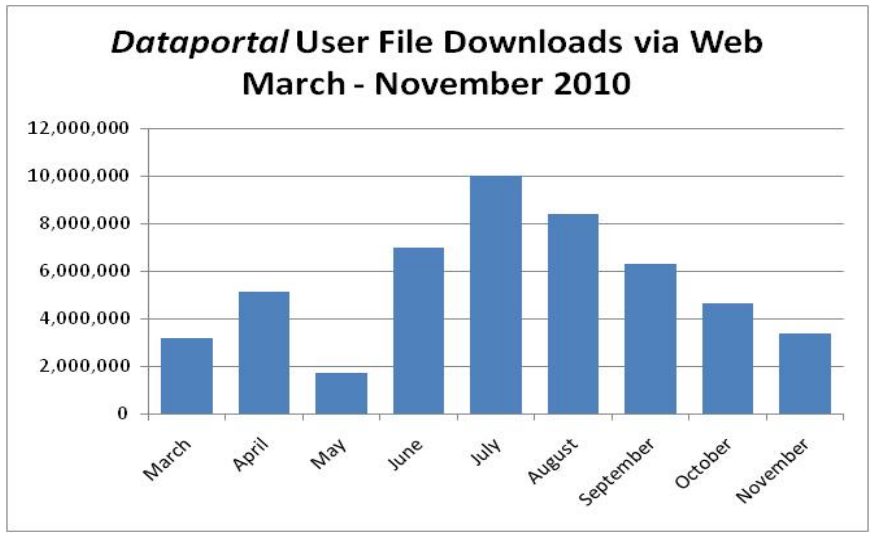
Archive Data Stored

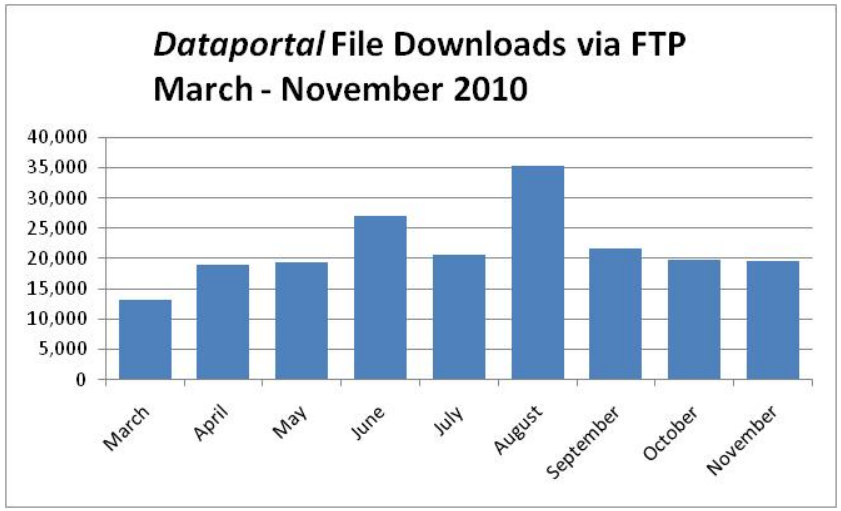


- NCCS pays an SGI license based on data stored (upgrading from15 to 20 PBs)
- Please remove any data you know you do not need

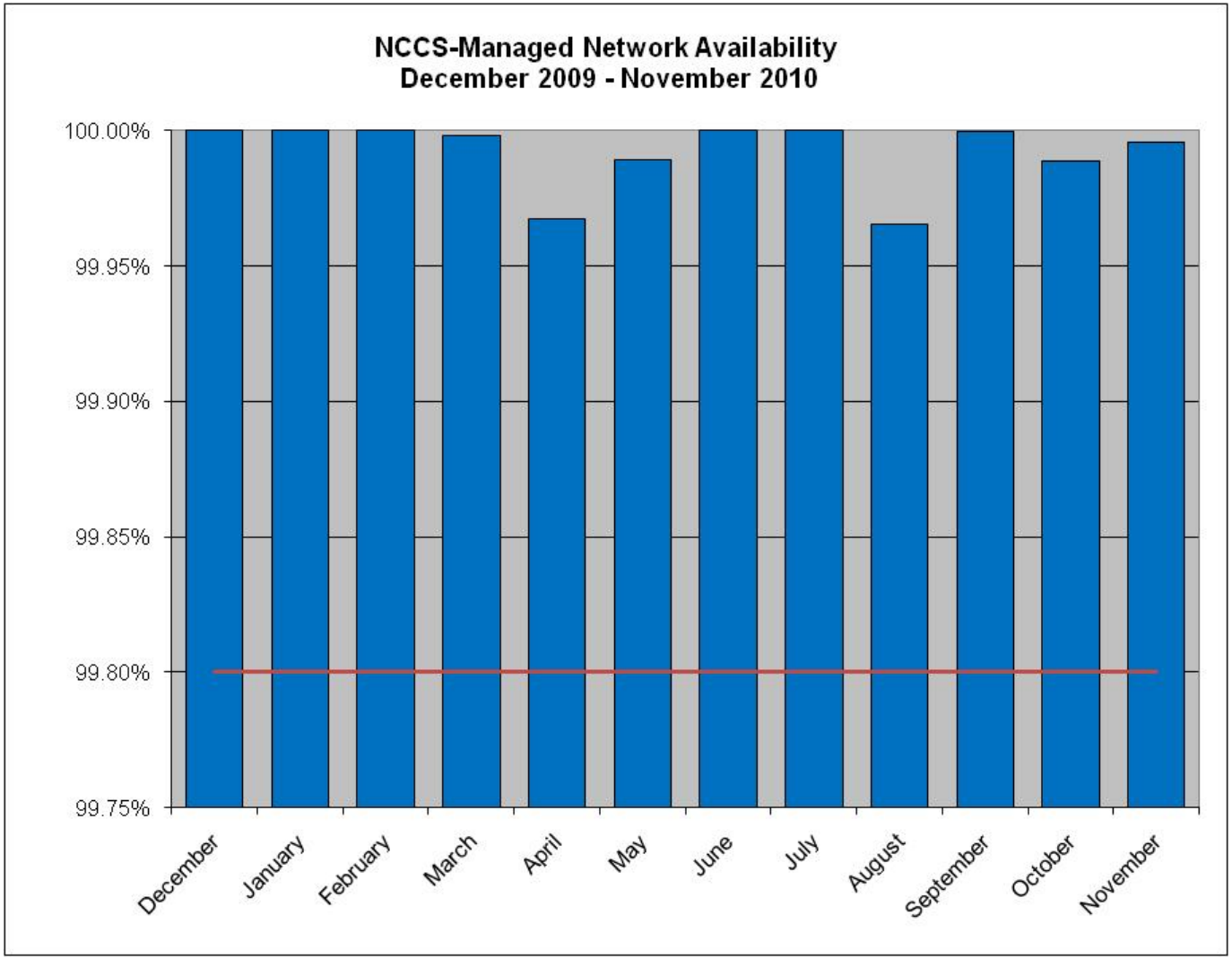
Dataportal File Downloads







NCCS Network Availability
Past 12 months

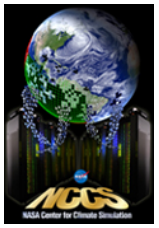


Archive Issues

1. Two CXFS bugs (server panic, CXFS failover does not complete)
   - Sending diagnostic information to SGI
   - SGI providing recommendations
2. Slow failover during actual problems (causes delayed resource availability)
   - Sending diagnostic information to SGI
   - SGI providing recommendations
3. System-level token lock (causes filesystem hangs)
   - Fix applied
   - Next step: Address new issue on Discover (see next slide), then reactivate NFS edge servers, then mount archive filesystems on Dali
4. Inappropriate failover when nothing was wrong
   - Automated monitoring errors
   - Implemented SGI-provided monitoring script changes
   - No further occurrences

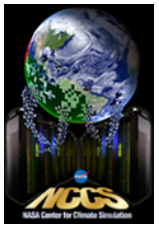
- General SGI support-related comments
  - Eager to resolve problems
  - Actual SGI code developers providing assistance
  - Support personnel very responsive

Discover Issues



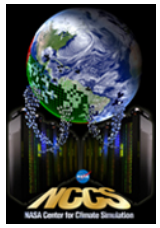
1. Intermittent slow I/O
   - Can lead to system hangs
   - GPFS designed for larger files, streaming I/O
   - NCCS is working with IBM, monitoring system and sending diagnostics to IBM as problems occur
   - Plan to implement partial fix 15 December
   - Please work with NCCS, SIVO if your application exhibits high file open/close activity or uses many small files
2. Jobs take longer to begin execution (insufficient capacity)
   - SCU7 will help
3. Users running nodes out of memory can cause GPFS hangs
   - Please work with NCCS, SIVO if your application runs nodes out of memory
4. New intermittent data corruption issue
   - Copy of data from archive to Discover via NFS sometimes results in null blocks

Upcoming Changes



- LDAP hardware upgrade
- Data Exploration Theater performance, functionality
- *Discover*
  - I/O performance changes
  - PBS v10
  - SLES 11
  - SCU7
- *DataPortal* – iRODS
- Mass Storage – Additional updates for increased stability
- Security – Firewall hardware, software, rule set changes
- Network – IP address changes

Agenda – December 7, 2010



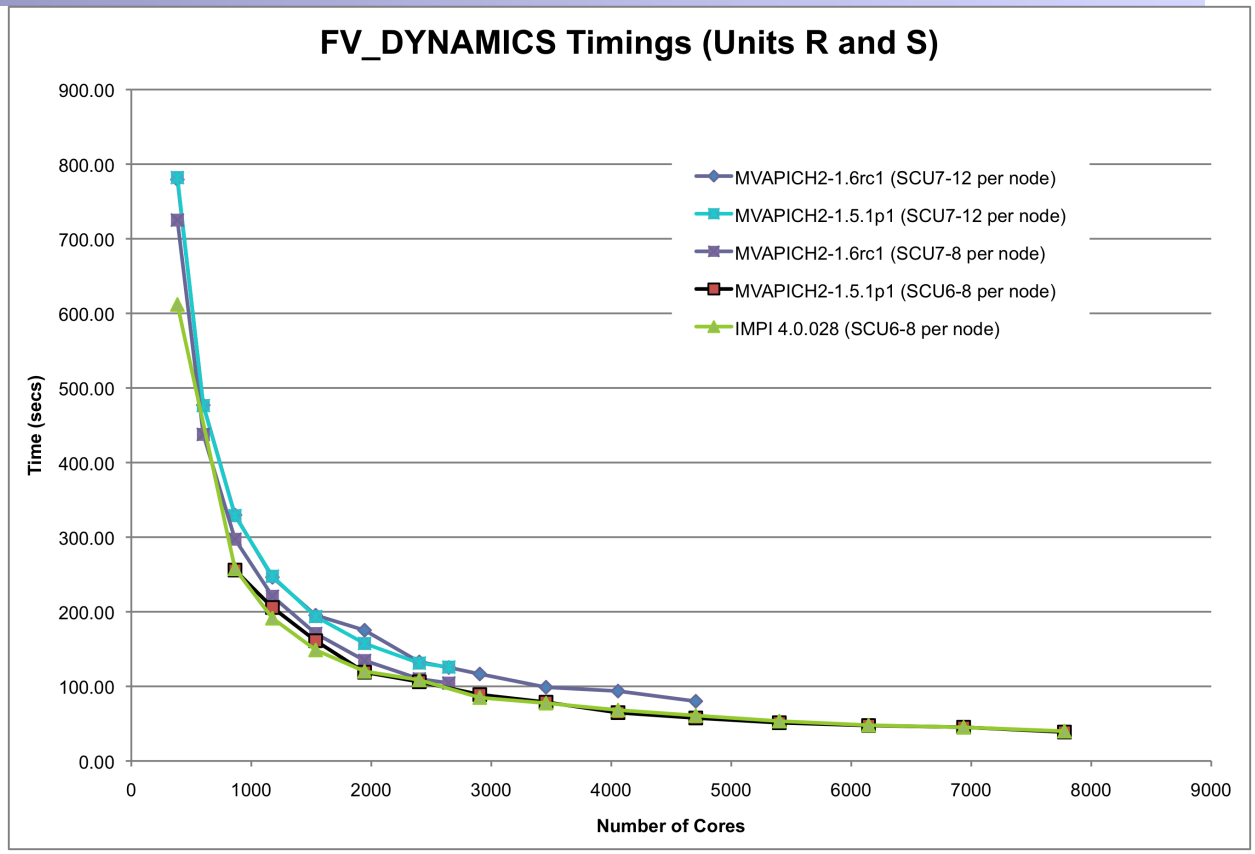
- Welcome & Introduction (Phil Webster, CISTO Chief)
- Current System Status (Fred Reitz, NCCS Operations Manager)
- **SCU7 and Other NCCS Systems Updates (Dan Duffy, NCCS Lead Architect)**
- Analysis Software Update (Tom Maxwell, NCCS Analysis Lead)
- User Services Update (Tyler Simon, NCCS User Services Group)
- Questions & Wrap-Up (Phil Webster)

SCU7 Updates

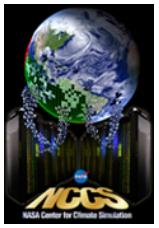
- Current status
  - System is physically installed, configured, and attached to the Discover cluster
  - Running SLES11 operating system (upgrade over the current version on Discover)
  - Will be running PBSv10 when general accessible by all users
  - System is not in production; it is in a dedicated pioneer state running large scale GEOS runs
- Current issues
  - Power and cooling concerns must be addressed prior to production

Performance Looks Good!

- Running 8 cores per node gives equivalent performance as on the Nehalem nodes (as expected)
- Running 12 cores per node results in about a 20% slowdown versus the same number of cores as the Nehalem nodes (also as expected from previous measurements)
- No difference in performance across different versions of MPI

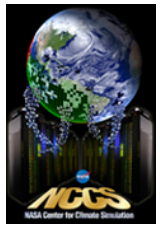


What's next with SCU7?



- NCCS and Dell are currently and very actively working on the power and cooling issues
- System will be maintained in a dedicated pioneer phase for now
  - Need to do this to minimize the disruption caused by changes to the system
- Target general availability by the end of January

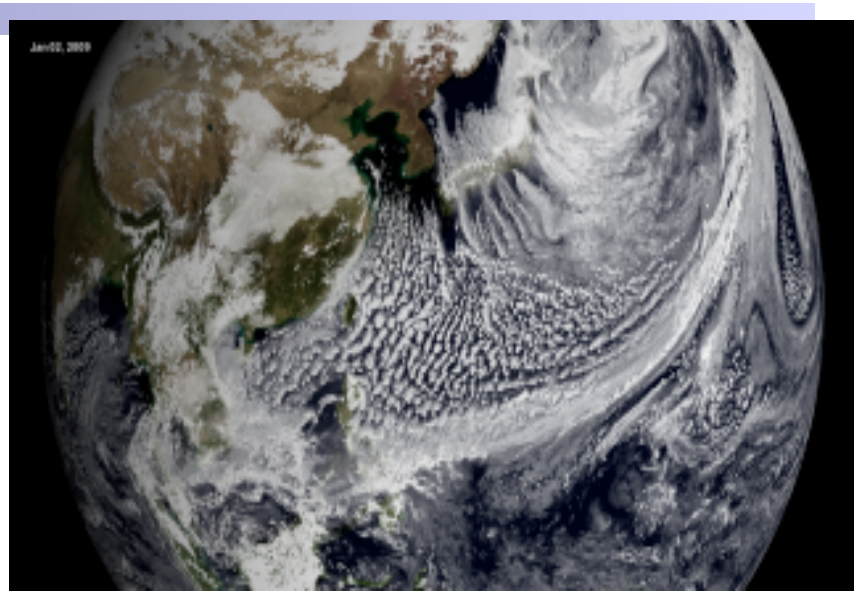
Agenda – December 7, 2010

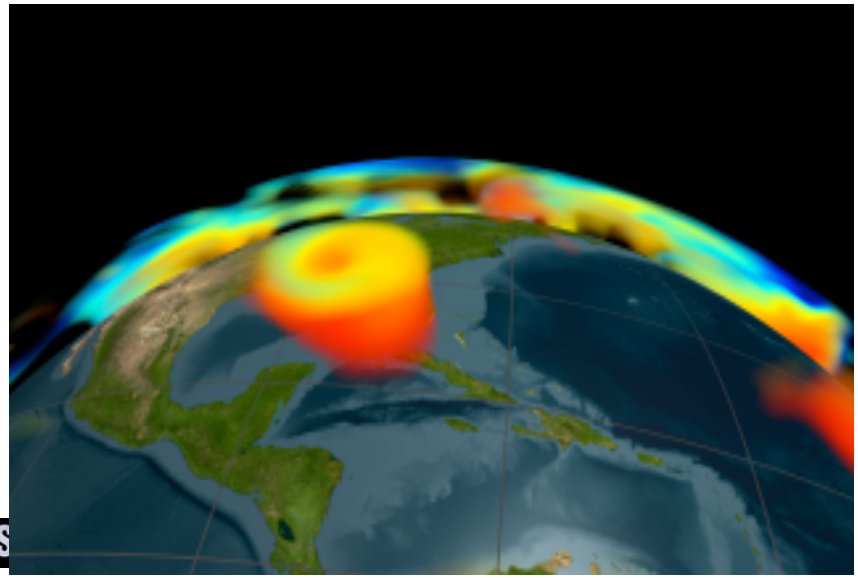


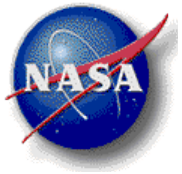
- Welcome & Introduction (Phil Webster, CISTO Chief)
- Current System Status (Fred Reitz, NCCS Operations Manager)
- SCU7 and Other NCCS Systems Updates (Dan Duffy, NCCS Lead Architect)
- **Analysis Software Update (Tom Maxwell, NCCS Analysis Lead)**
- User Services Update (Tyler Simon, NCCS User Services Group)
- Questions & Wrap-Up (Phil Webster)

Ultra-Analysis Requirements

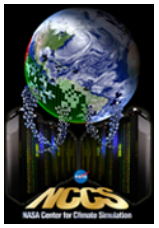
- Parallel streaming analysis pipelines.
  - Data parallelism.
  - Task parallelism.
- Parallel IO.
- Remote interactive execution.
- Advanced visualization.
- Provenance capture.
- Interfaces for scientists.
  - Workflow construction tools.
  - Visualization interfaces.







BER Earth System Modeling Proposals

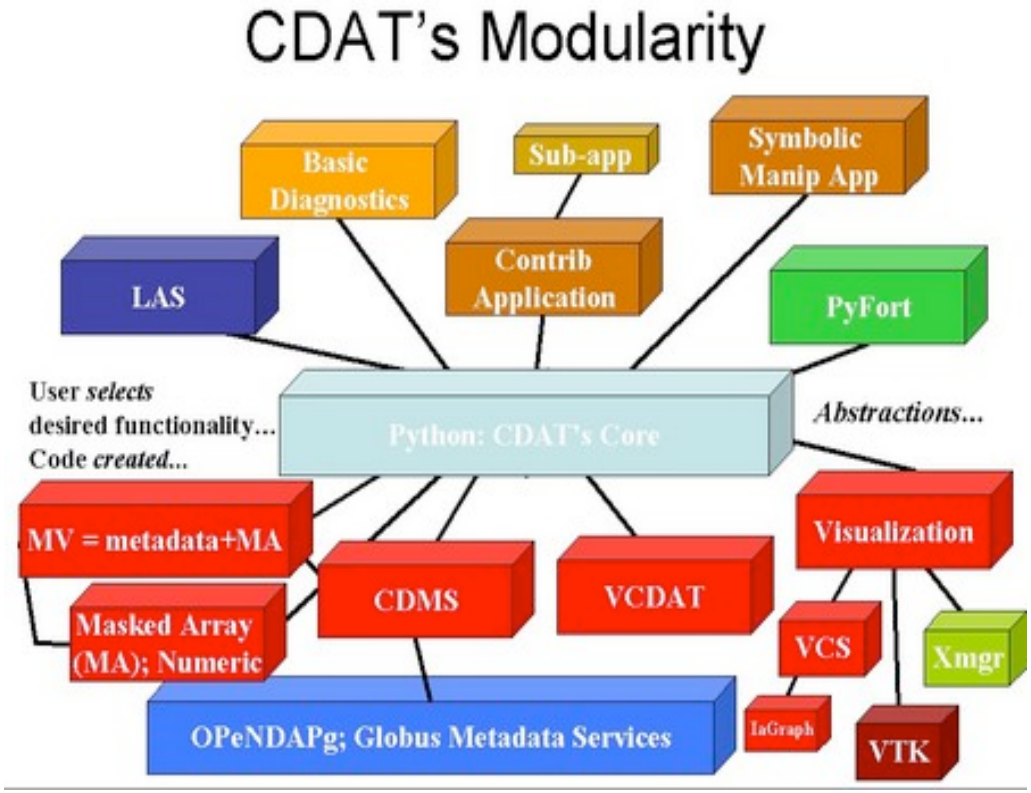


Advanced Scientific Visualization of Ultra-Large Datasets

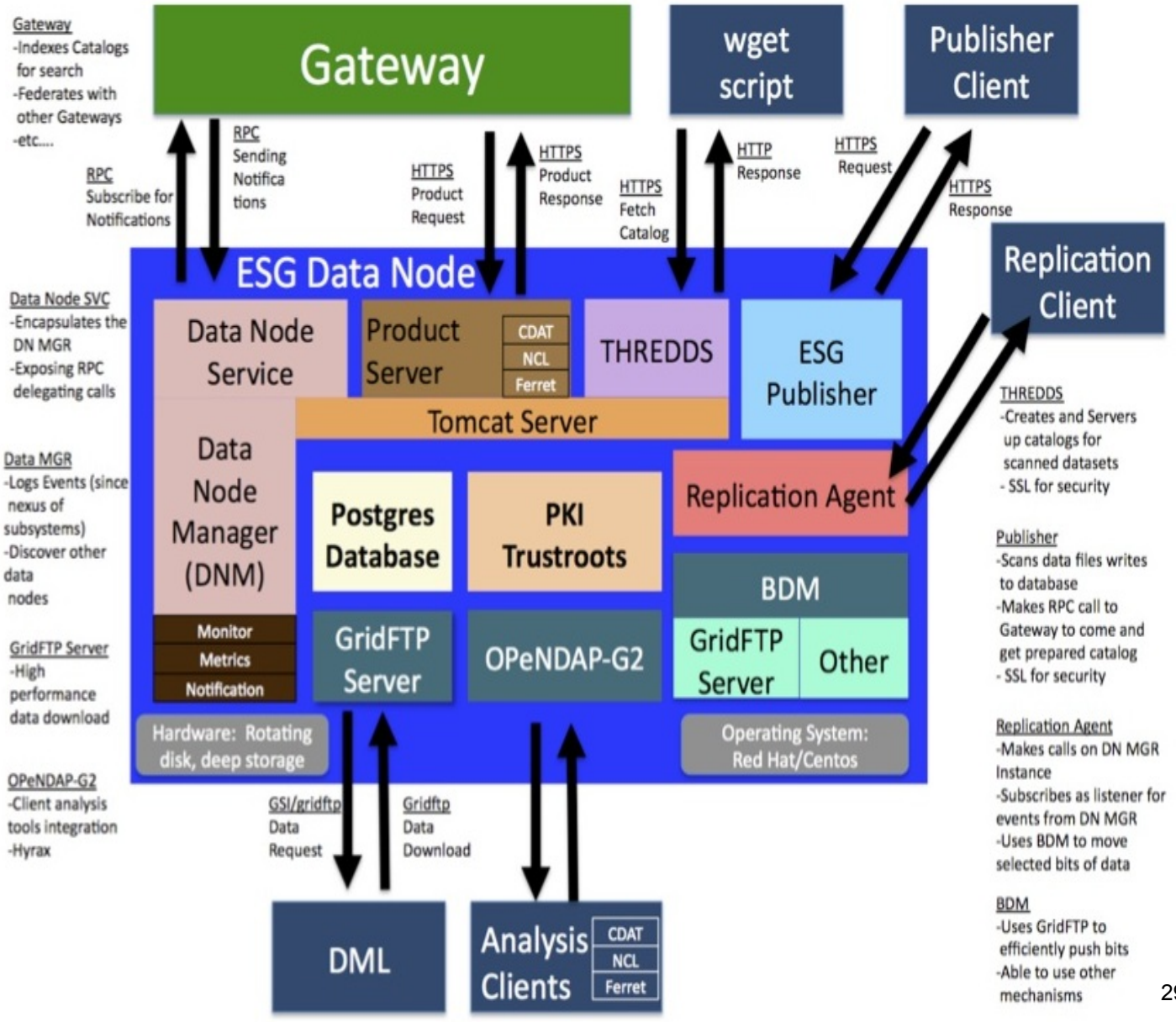
Topic	PI	Location
UV-CDAT development	Dean Williams	LLNL, PCMDI
Visit development	Wes Bethel	LBNL
NCO parallelization	Robert Jacob	ANL

Climate Data Analysis Toolkit

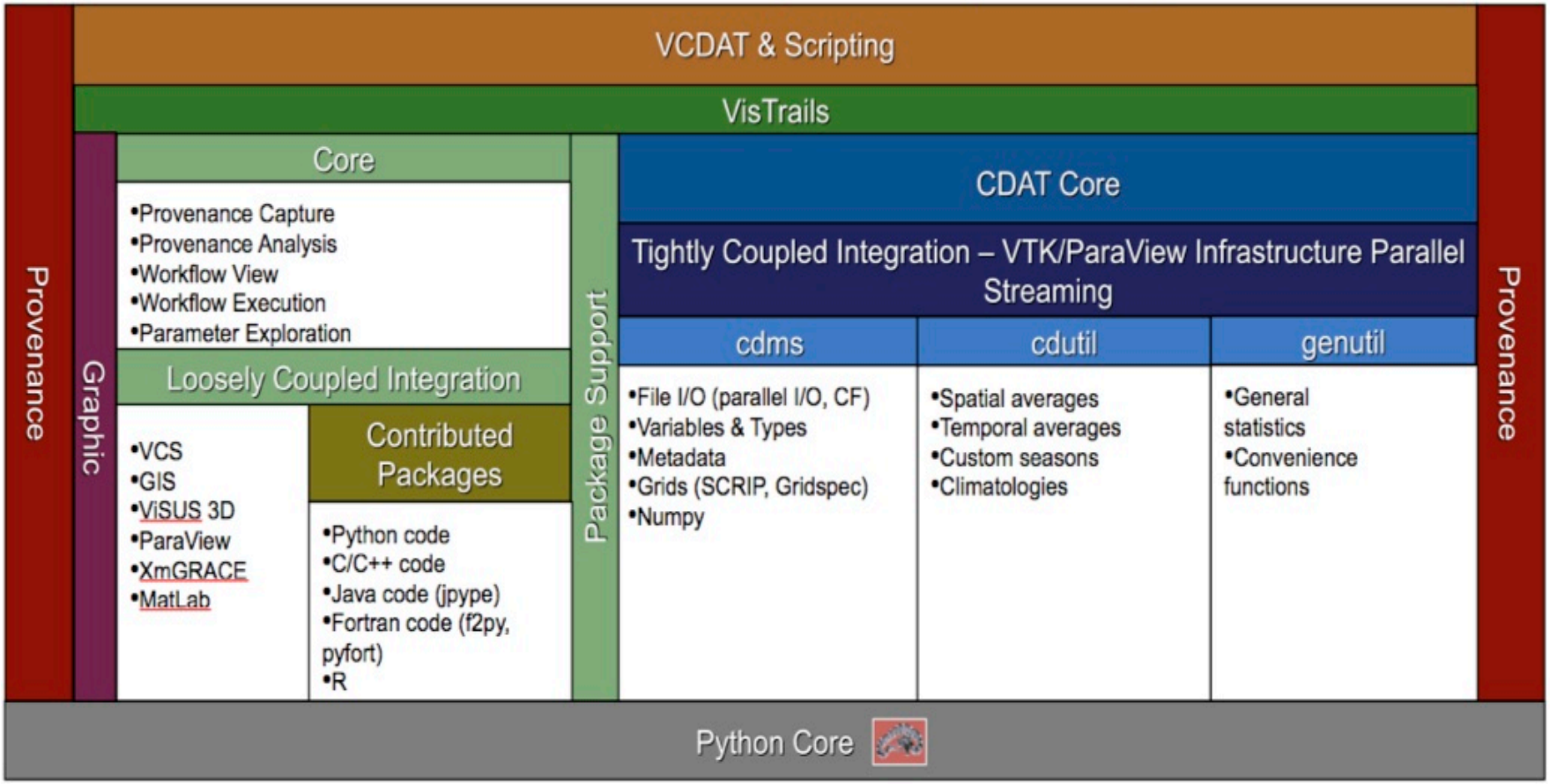
- Integrated environment for data processing, visualization, & analysis.
- Integrates numerous software modules in python shell.
- Open source with a large diverse set of contributors.
- Analysis environment for ESG developed @ LLNL.

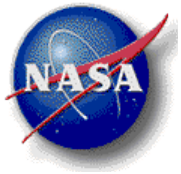


Earth System Grid



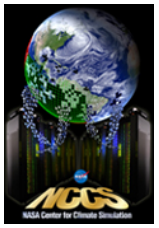
UV-CDAT Architecture





ParaView

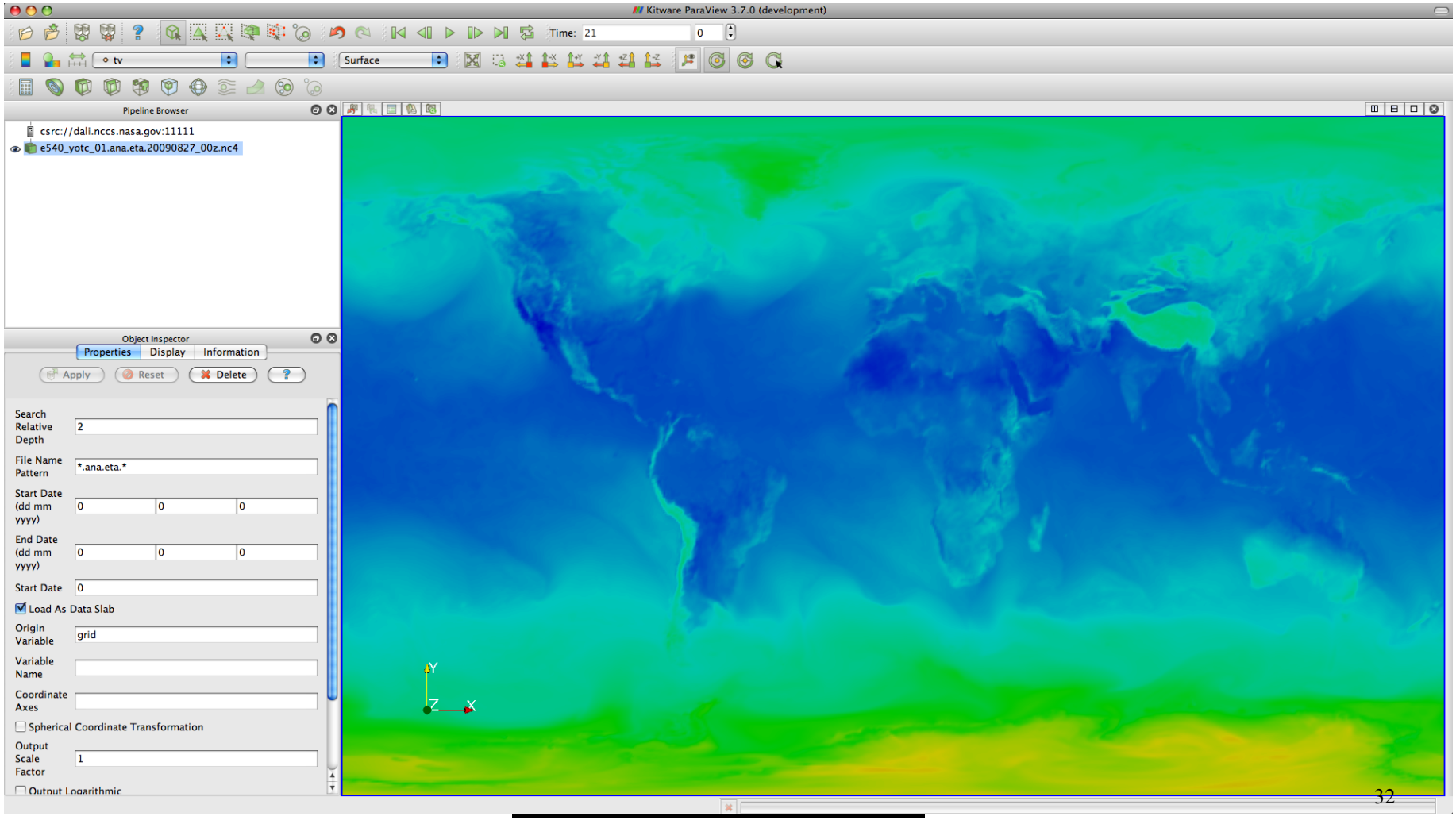
Parallel Visualization Application



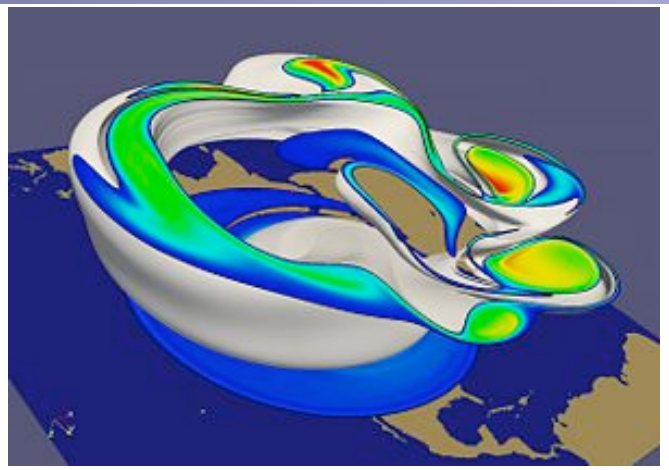
- Open-source, multi-platform visualization application.
  - Developed by Kitware, Inc. (authors of VTK).
- Designed to process large data sets.
- Built on parallel VTK.
- Client-server architecture:
  - Client: Qt based desktop application.
  - Data Server: MPI based parallel application.
- Parallel streaming IO & pipeline for data processing.
- Large library of existing filters.
- Highly extensible using plugins.
- No existing climate-specific tools or algorithms.
- Data Server being integrated into ESG.

ParaView Client

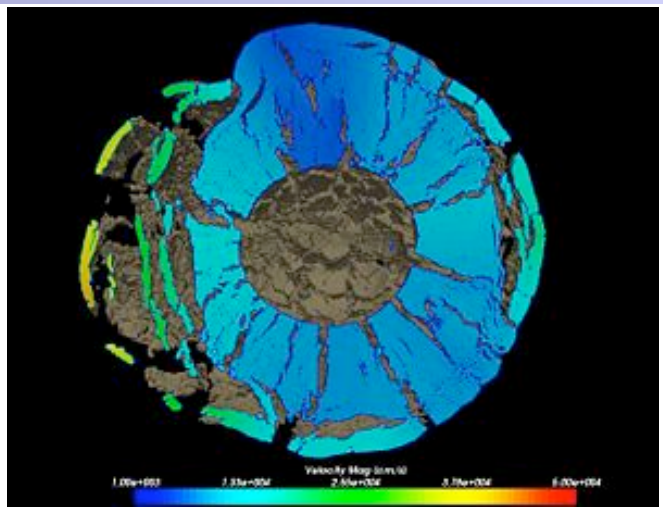
Qt desktop application: Controls data access, processing, analysis, and viz.



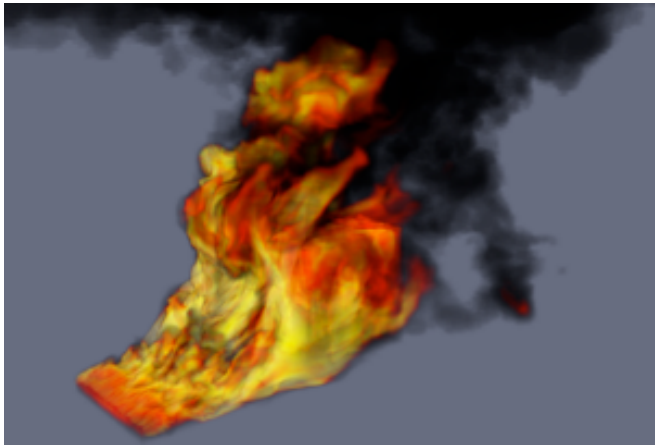
ParaView Applications



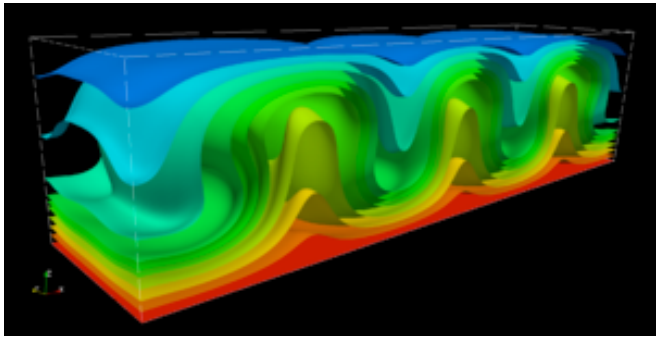
Polar Vortex Breakdown Simulation



Golevka Asteroid Explosion Simulation



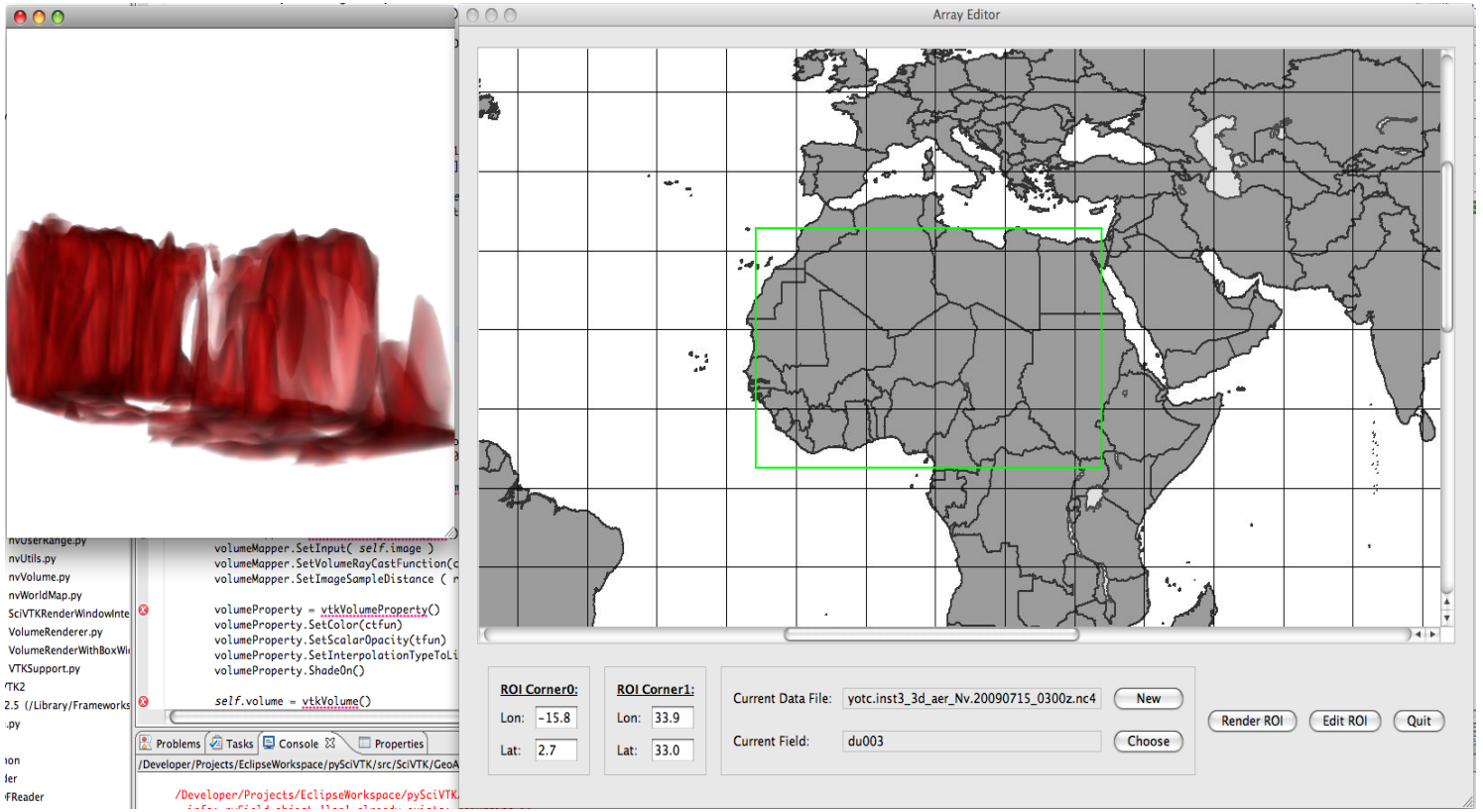
Cross Wind Fire Simulation



3D Rayleigh-Benard problem

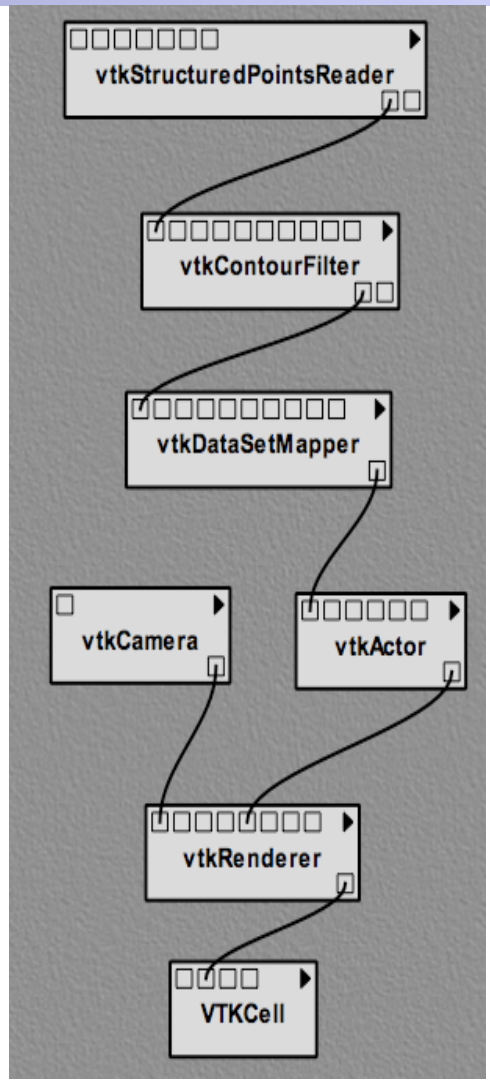
Python VTK tools

- Integrate VTK 3D visualization into GrADS and CDAT for climate science applications.
- Develop high level python/Qt interfaces to simplify common scientific visualization tasks.



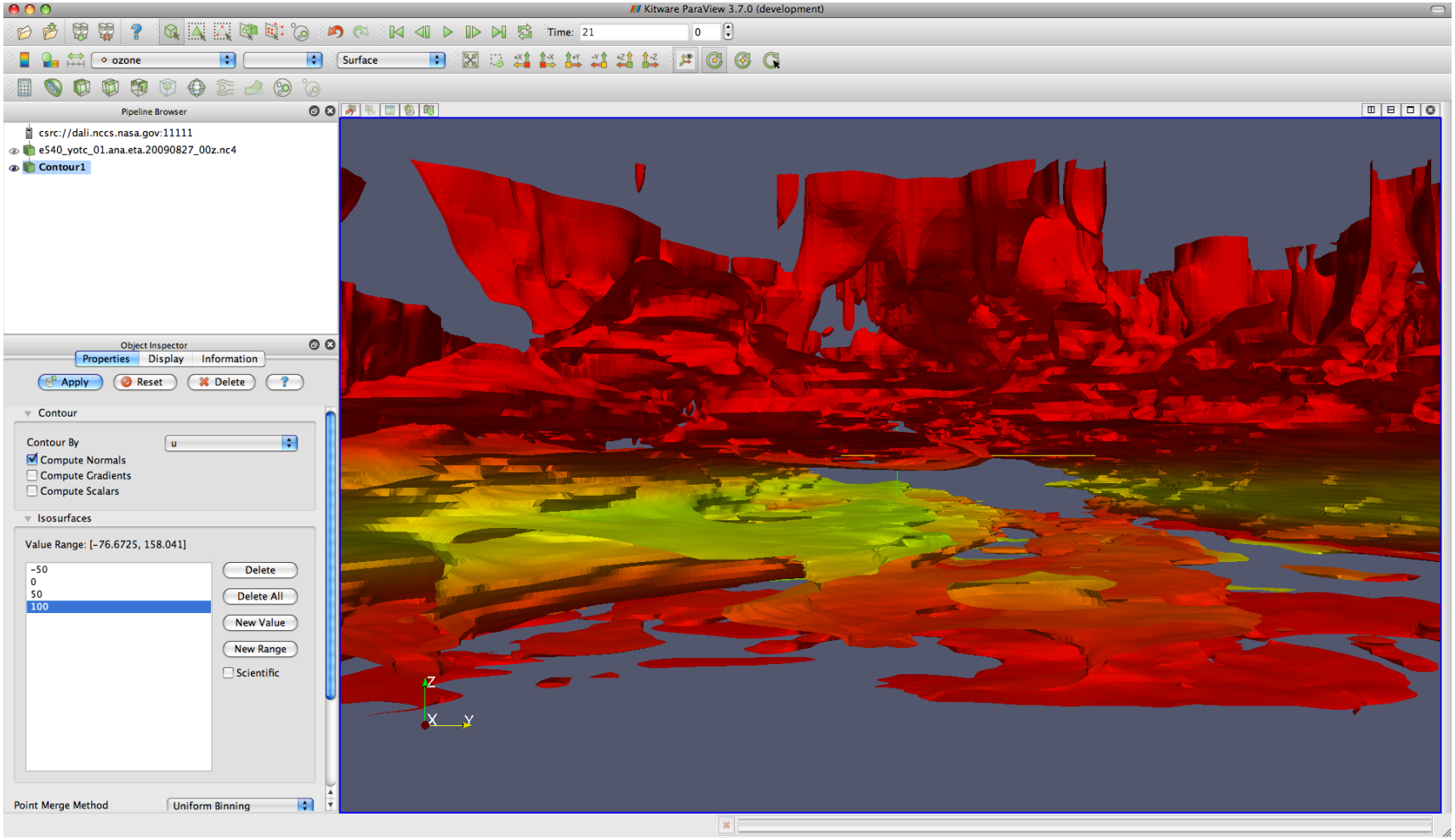
Analysis Workflows

```
1  import vtk
2
3  data = vtk.vtkStructuredPointsReader()
4  data.SetFileName("../examples/data/head.120.vtk")
5
6  contour = vtk.vtkContourFilter()
7  contour.SetInput(0,data.GetOutput())
8  contour.SetValue(0, 67)
9
10 mapper = vtk.vtkPolyDataMapper()
11 mapper.SetInput(contour.GetOutput())
12 mapper.ScalarVisibilityOff()
13
14 actor = vtk.vtkActor()
15 actor.SetMapper(mapper)
16
17 cam = vtk.vtkCamera()
18 cam.SetViewUp(0,0,-1)
19 cam.SetPosition(745,-453,369)
20 cam.SetFocalPoint(135,135,150)
21 cam.ComputeViewPlaneNormal()
22
23 ren = vtk.vtkRenderer()
24 ren.AddActor(actor)
25 ren.SetActiveCamera(cam)
26 ren.ResetCamera()
27
28 renwin = vtk.vtkRenderWindow()
29 renwin.AddRenderer(ren)
30
31 style = vtk.vtkInteractorStyleTrackballCamera()
32 iren = vtk.vtkRenderWindowInteractor()
33 iren.SetRenderWindow(renwin)
34 iren.SetInteractorStyle(style)
35 iren.Initialize()
36 iren.Start()
```

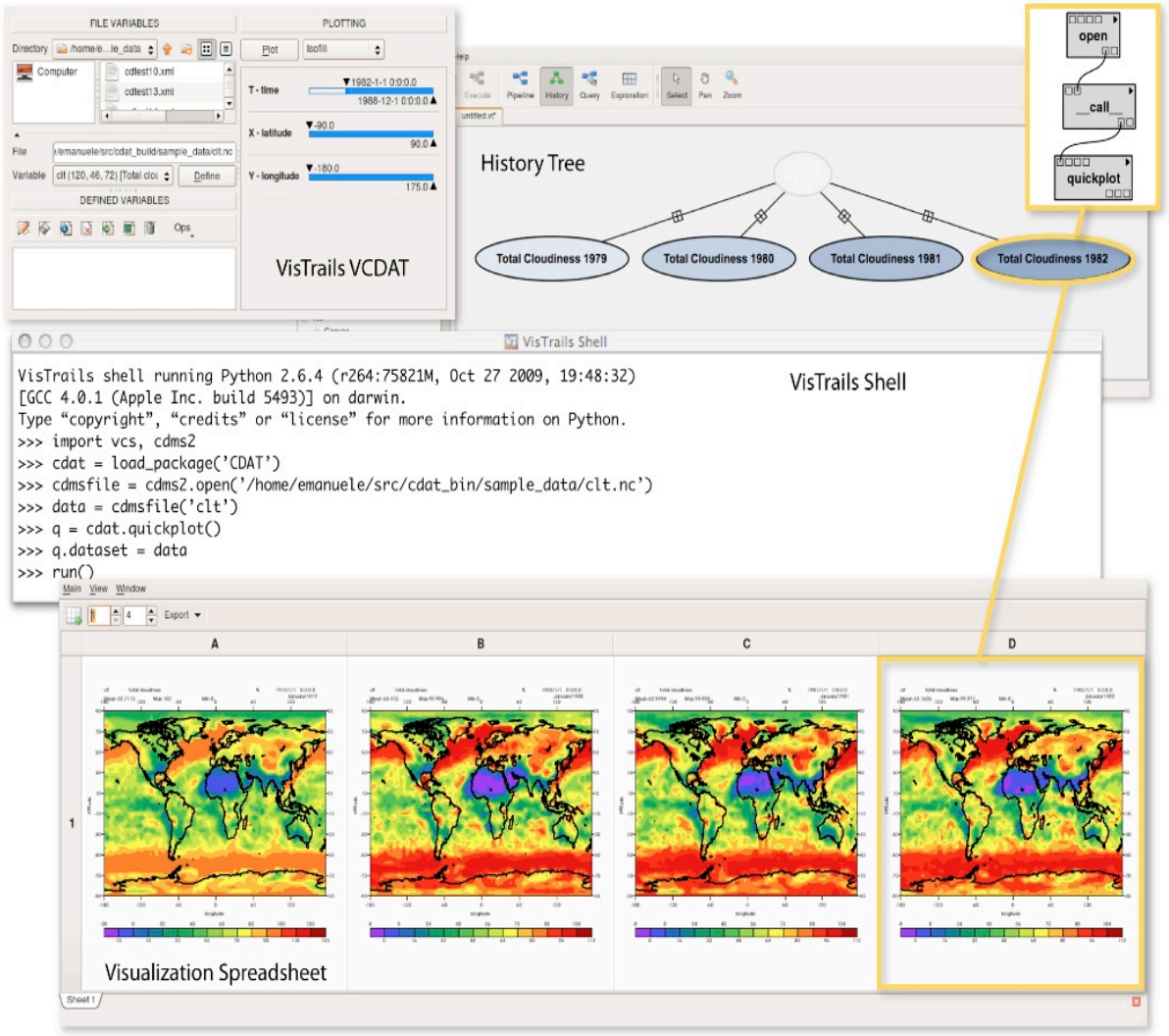


Analysis Workflow Configuration

Configure a parallel streaming pipeline for data analysis

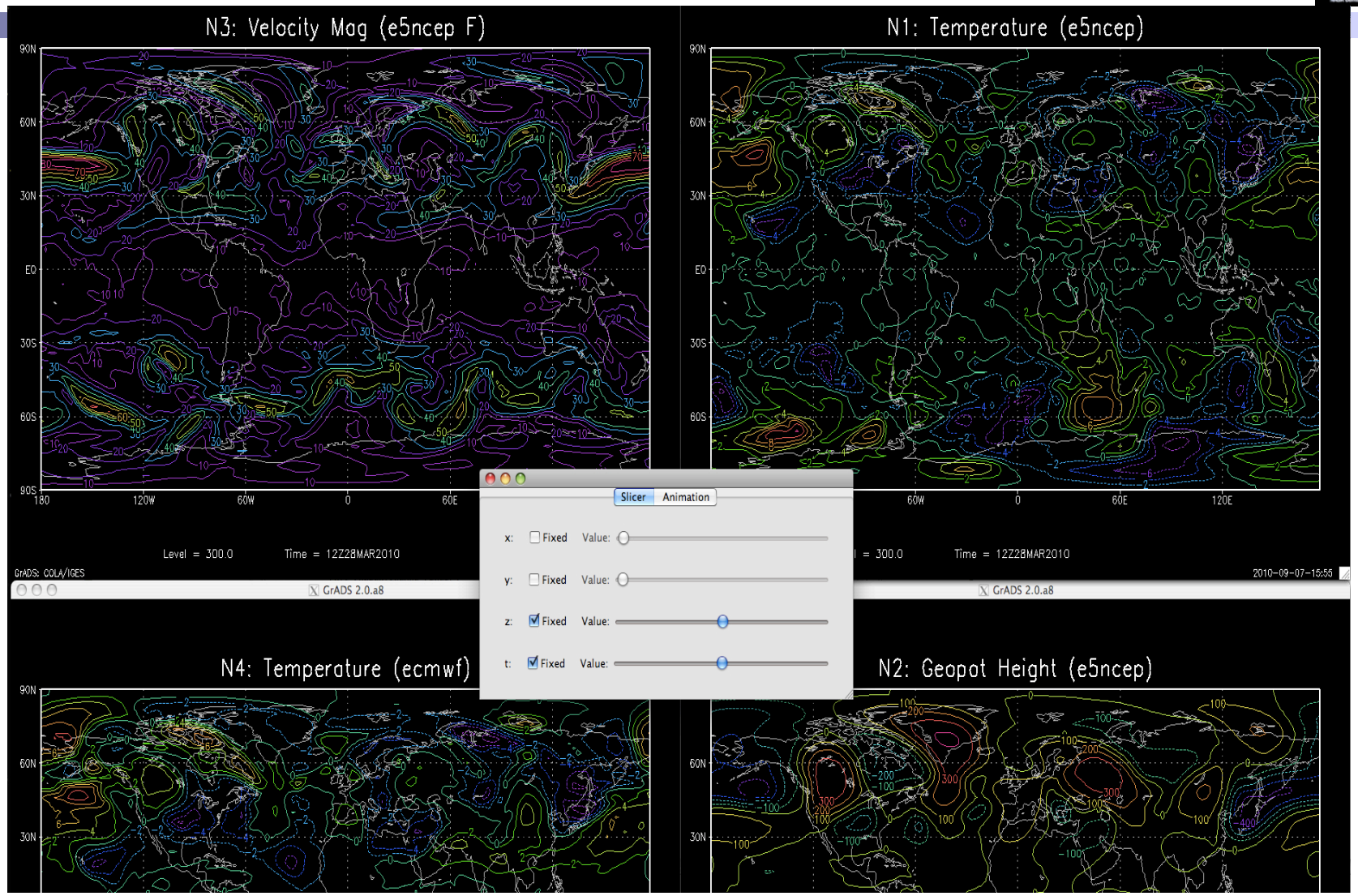


UV-CDAT Interface



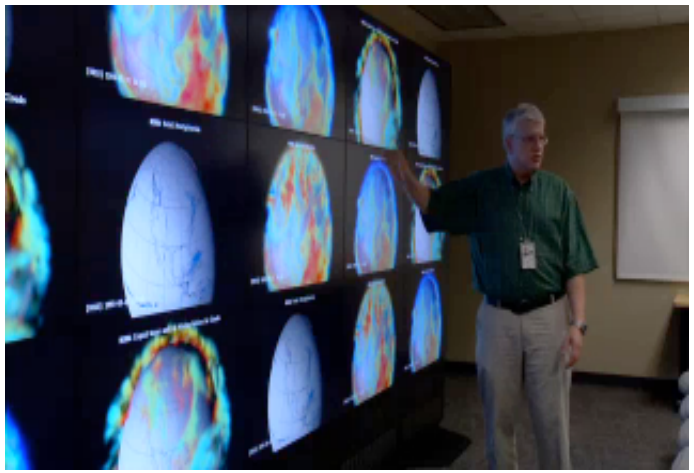
NCDAS

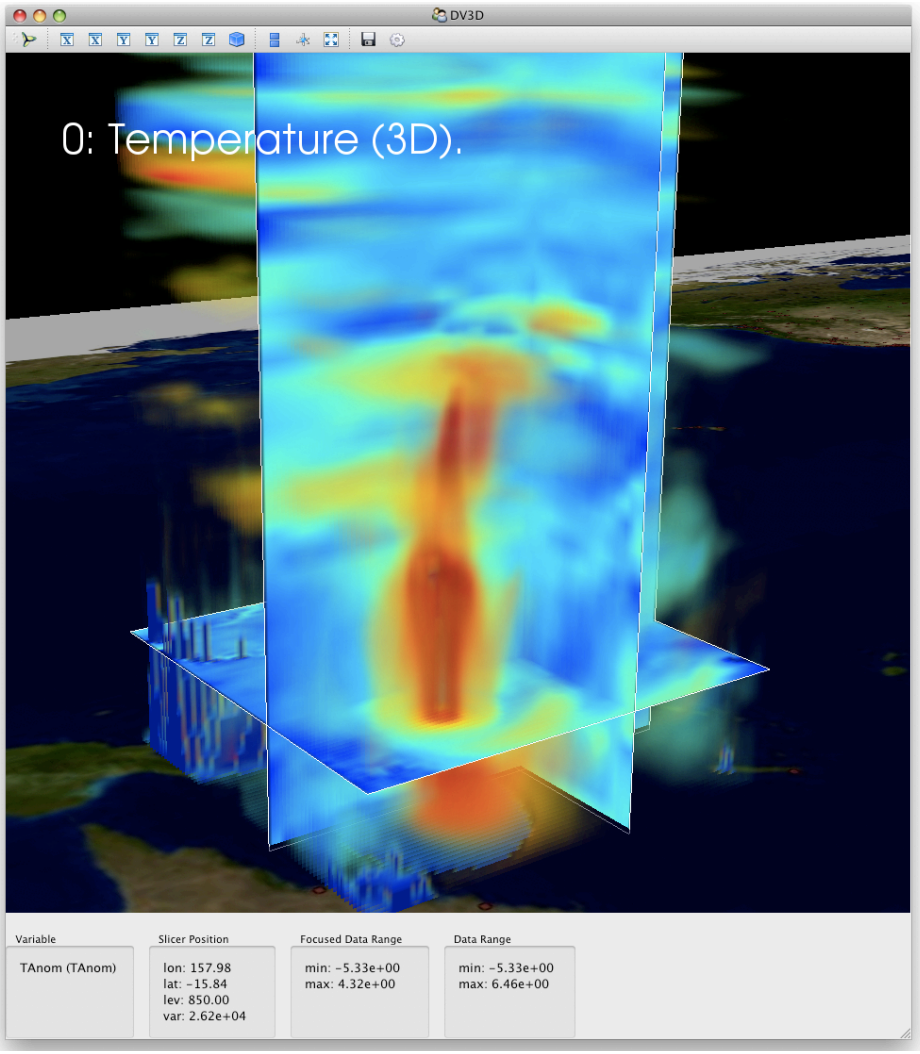
Task parallel interactive data analysis
Multiple views of dataset



DV3D

- Visualization & analysis application.
- Built in python on MayaVi / VTK.
- Tailored for climate scientists.
- Simple intuitive interface.
- Integrated analysis frameworks:
  - pyGrads, UVCDAT.
- Multiple views via hyperwall.
- Interactive 3D data browsing.
- Available on dali, DET, ford1

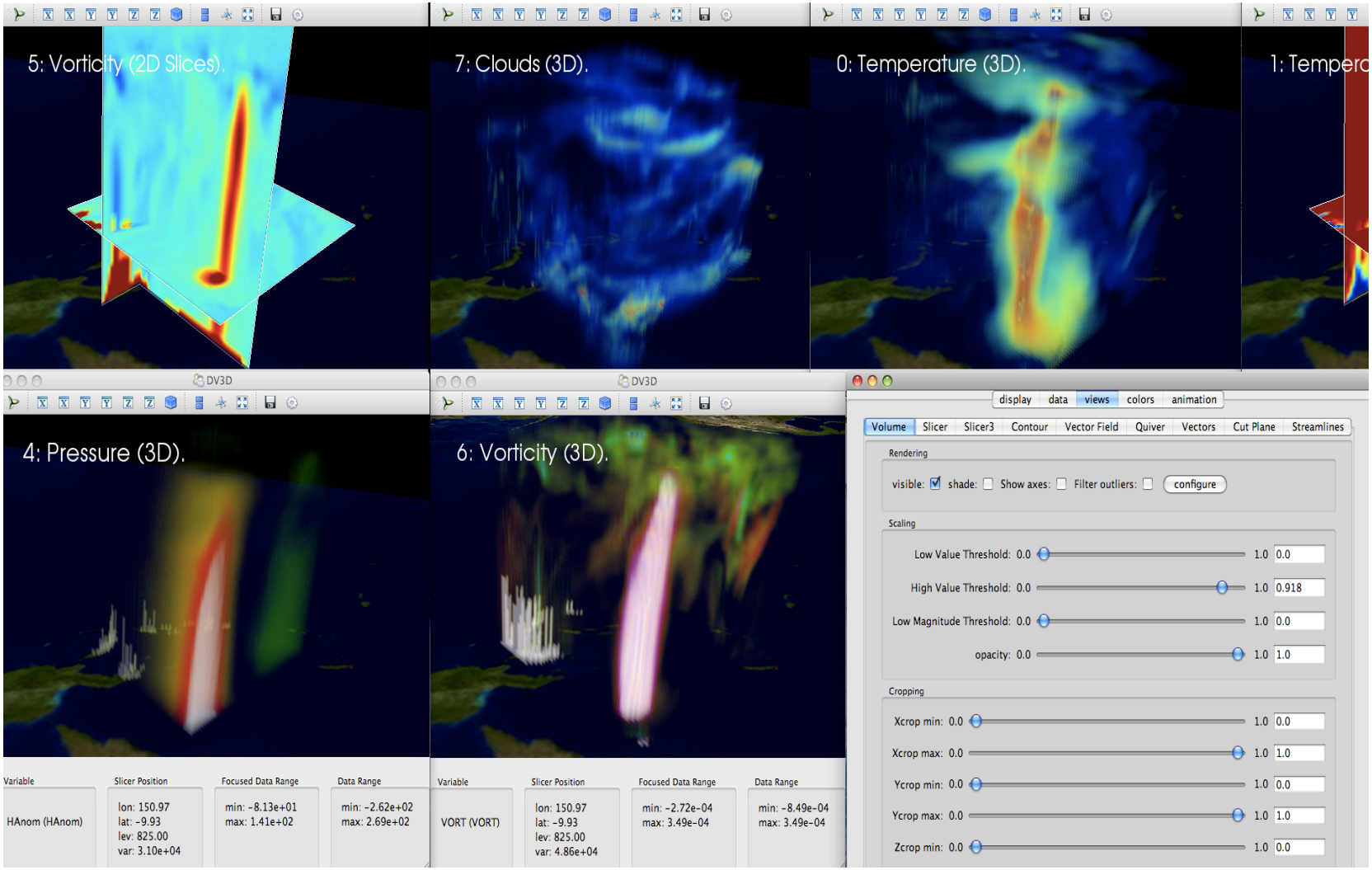




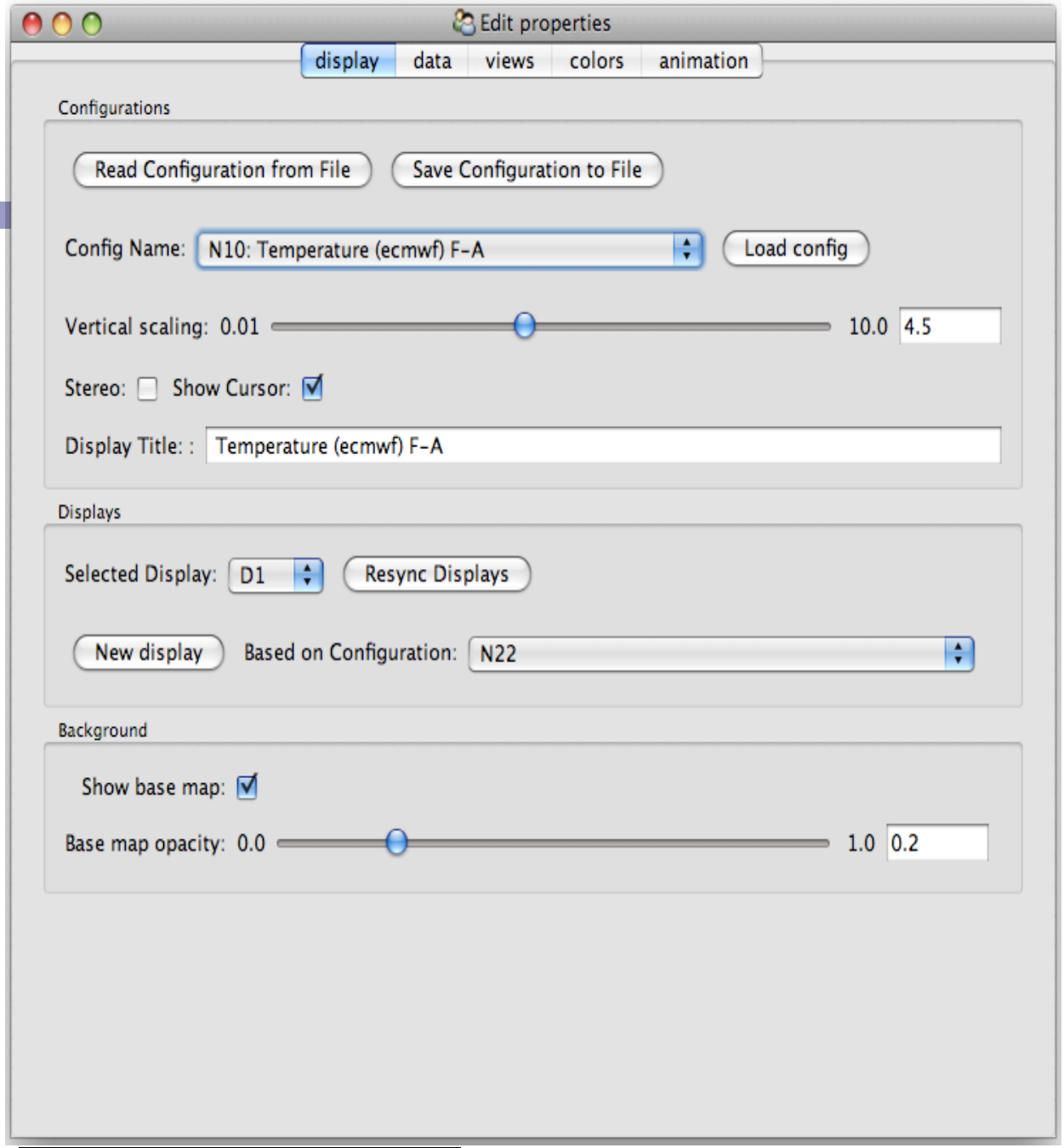
DV3D

Interactive 3D visualization of simulation data

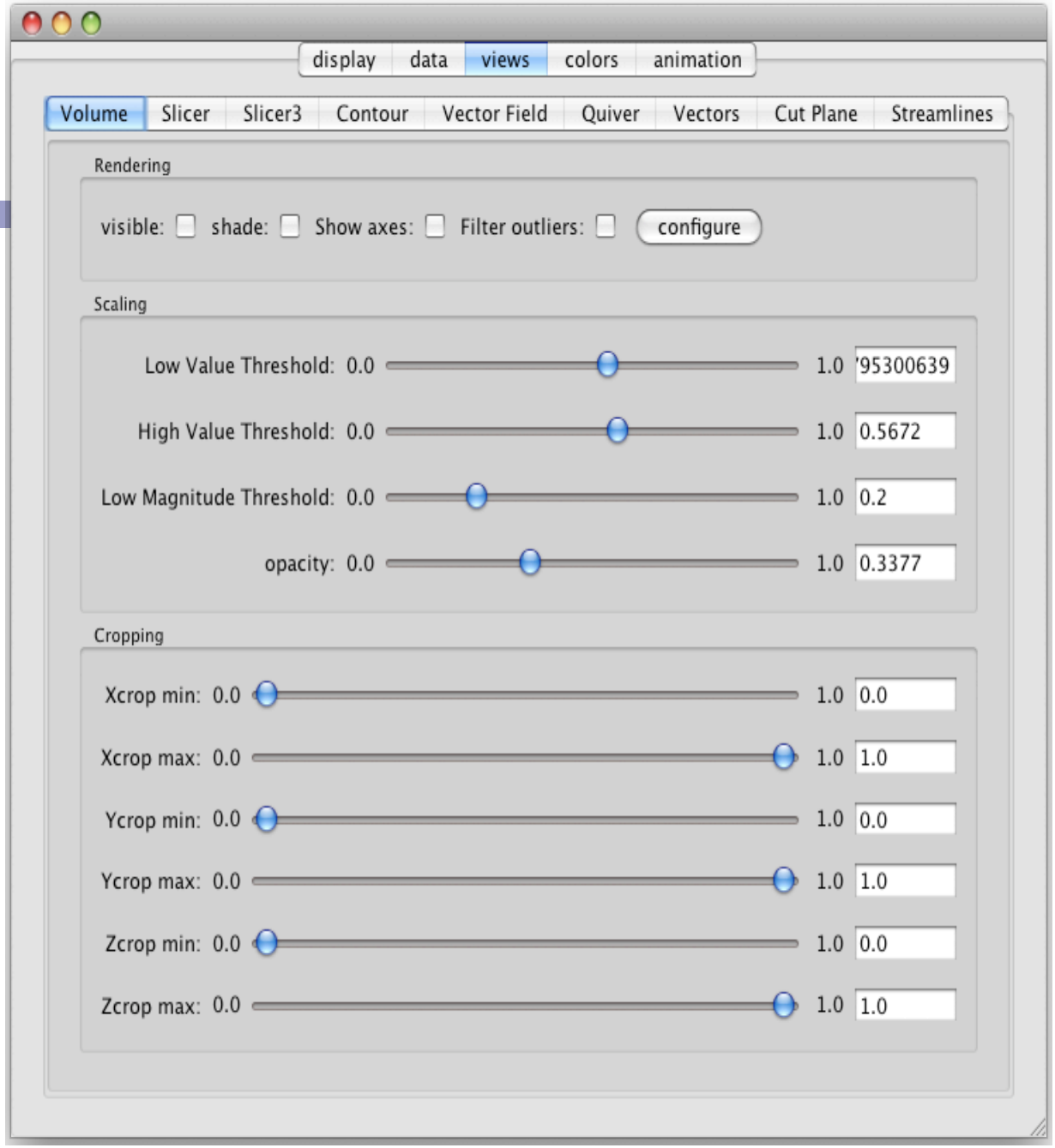
http://portal.nccs.nasa.gov/DV3D/



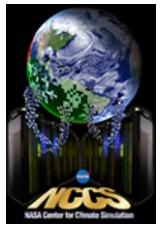
DV3D Display Panel



DV3D Volume Rendering Panel

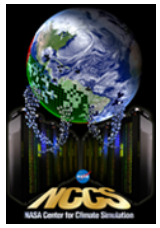


Analysis Tools Under Development



- Integrate DV3D into UVCDAT / Vistrails / ParaView / ESG
- Remote analysis and visualization of climate data on dali
  - Clients: hyperwall, desktop, SC-11
- In-situ analysis and real-time visualization of GEOS-5 runs
- 3D stereo visualization of climate data in DET
- UVCDAT Use Cases:
  - Hurricane Tracking with high resolution GEOS-5 data
- Seeking Scientific Use Cases

Agenda – December 7, 2010



- Welcome & Introduction (Phil Webster, CISTO Chief)
- Current System Status (Fred Reitz, NCCS Operations Manager)
- SCU7 and Other NCCS Systems Updates (Dan Duffy, NCCS Lead Architect)
- Analysis Software Update (Tom Maxwell, NCCS Analysis Lead)
- **User Services Update (Tyler Simon, NCCS User Services Group)**
- Questions & Wrap-Up (Phil Webster)

User Services

- Intel MPI 4.0 and MVAPICH2 1.6?
- Discover node profiles:

Processor Type	Cores/node Frequency (GHz) clocks/flop	Memory per node (GB)	Total cores	PBS
Dempsey	4 @ 3.2 (2)	4, **1GB per core**	520	Proc=demp
Woodcrest	4 @ 2.66 (4)	4, **1 GB per core**	2,064	Proc=wood
Harpertown	8 @ 2.5 (4)	16, **2 GB per core**	4,128	Proc=harp
Nehalem	8 @ 2.8 (4)	24, **3GB per core**	8,256	Proc=neha
Westmere	12 @ 2.8 (4)	24, **2GB per core**	14,400	Proc=west?

```
#PBS -l select=64:ncpus=8:proc=neha
```

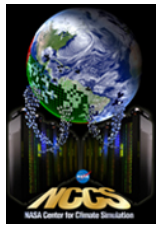
User Services

- Matlab Licenses?
- Expansion Factor = $\frac{\text{Wait time + Job Runtime}}{\text{Job Runtime}}$
  - Minimum Expansion factor is 1, when wait time = 0.

- Future Scheduling enhancements?
  - **Time to solution** = Wait time + Job Runtime
  - Does anyone have a problem with this?
  - You want to know the TTS not the Expansion factor right?
- Naturally we want to focus on providing and minimizing Time to Solution for every job.

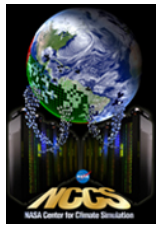
NCCS Job Monitor

Agenda – December 7, 2010



- Welcome & Introduction (Phil Webster, CISTO Chief)
- Current System Status (Fred Reitz, NCCS Operations Manager)
- SCU7 and Other NCCS Systems Updates (Dan Duffy, NCCS Lead Architect)
- Analysis Software Update (Tom Maxwell, NCCS Analysis Lead)
- User Services Update (Tyler Simon, NCCS User Services Group)
- **Questions & Wrap-Up (Phil Webster)**

Contact Information



NCCS User Services:

support@nccs.nasa.gov

301-286-9120

https://www.nccs.nasa.gov

Thank you